B4

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 August 2002 (01.08.2002)

PCT

(10) International Publication Number
WO 02/059312 A2

(51) International Patent Classification[7]: C12N 15/12, C07K 14/705, C12N 5/10, C07K 16/18, C12Q 1/68, A61K 31/00, 33/00, 38/17, 39/395, G01N 33/50

(21) International Application Number: PCT/US01/49206

(22) International Filing Date:
18 December 2001 (18.12.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/256,542	18 December 2000 (18.12.2000)	US
60/259,604	22 December 2000 (22.12.2000)	US
60/260,101	5 January 2001 (05.01.2001)	US

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **DUGGAN, Brendan, M.** [AU/US]; 243 Buena Vista Avenue #306, Sunnyvale, CA 94086 (US). **XU, Yuming** [US/US]; 1739 Walnut Drive, Mountain View, CA 94040 (US). **LEE, Ernestine, A.** [US/US]; 624 Kains Street, Albany, CA 94706 (US). **LEE, Sally** [US/US]; 825 East Evelyn, #425, Sunnyvale, CA 94086 (US). **LU, Dyung, Aina, M.** [US/US]; 233 Goy Drive, San Jose, CA 95123 (US). **WARREN, Bridget, A.** [US/US]; 10130 Parwood Drive #2, Cupertino, CA 95014 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **GIETZEN, Kimberly, J.** [US/US]; 691 Los Huecos Drive, San Jose, CA 95123 (US). **HONCHELL, Cynthia, D.** [US/US]; 400 Laurel Street #203, San Carlos, CA 94070 (US). **BURFORD, Neil** [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **TANG, Tom, Y.** [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **HILLMAN, Jennifer, L.** [US/US]; 230 Monroe Drive #17, Montain View, CA 94040 (US). **GANDHI, Ameena, R.** [US/US]; 705 5th Avenue, San Francisco, CA 94118 (US). **KALLICK, Deborah, A.** [US/US]; 900 Olive Street, Menlo Park, CA 94025 (US). **BANDMAN, Olga** [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). **GRAUL, Richard, C.** [—/US]; 682-29th Avenue, San Francisco, CA 94121 (US). **WALIA, Narinder, K.** [US/US]; 890 Davis Street #205, San Leandro, CA 94577 (US). **LU, Yan** [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **YAO, Monique, G.** [US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US). **LAL, Preeti, G.** [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/059312 A2

(54) Title: CELL ADHESION PROTEINS

(57) Abstract: The invention provides human cell adhesion proteins (CADHP) and polynucleotides which identify and encode CADHP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of CADHP.

CELL ADHESION PROTEINS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of cell adhesion proteins and to the use of these sequences in the diagnosis, treatment, and prevention of immune system disorders, neurological disorders, developmental disorders, and cell proliferative disorders, including cancer, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of cell adhesion proteins.

BACKGROUND OF THE INVENTION

The surface of a cell is rich in transmembrane proteoglycans, glycoproteins, glycolipids, and receptors. These macromolecules mediate adhesion with other cells and with components of the extracellular matrix (ECM). The interaction of the cell with its surroundings profoundly influences cell shape, strength, flexibility, motility, and adhesion. These dynamic properties are intimately associated with signal transduction pathways controlling cell proliferation and differentiation, tissue construction, and embryonic development. Families of cell adhesion molecules include the cadherins, integrins, lectins, neural cell adhesion proteins, and some members of the proline-rich proteins.

Cadherins comprise a family of calcium-dependent glycoproteins that function in mediating cell-cell adhesion in virtually all solid tissues of multicellular organisms. These proteins share multiple repeats of a cadherin-specific motif, and the repeats form the folding units of the cadherin extracellular domain. Cadherin molecules cooperate to form focal contacts, or adhesion plaques, between adjacent epithelial cells. The cadherin family includes the classical cadherins and protocadherins. Classical cadherins include the E-cadherin, N-cadherin, and P-cadherin subfamilies. E-cadherin is present on many types of epithelial cells and is especially important for embryonic development. N-cadherin is present on nerve, muscle, and lens cells and is also critical for embryonic development. P-cadherin is present on cells of the placenta and epidermis. Recent studies report that protocadherins are involved in a variety of cell-cell interactions (Suzuki, S.T. (1996) J. Cell Sci. 109:2609-2611). The intracellular anchorage of cadherins is regulated by their dynamic association with catenins, a family of cytoplasmic signal transduction proteins associated with the actin cytoskeleton. The anchorage of cadherins to the actin cytoskeleton appears to be regulated by protein tyrosine phosphorylation, and the cadherins are the target of phosphorylation-induced junctional disassembly (Aberle, H. et al. (1996) J. Cell. Biochem. 61:514-523).

Integrins are ubiquitous transmembrane adhesion molecules that link the ECM to the internal

cytoskeleton. Integrins are composed of two noncovalently associated transmembrane glycoprotein subunits called α and β. Integrins function as receptors that play a role in signal transduction. For example, binding of integrin to its extracellular ligand may stimulate changes in intracellular calcium levels or protein kinase activity (Sjaastad, M.D. and Nelson, W.J. (1997) BioEssays 19:47-55). At least ten cell surface receptors of the integrin family recognize the ECM component fibronectin, which is involved in many different biological processes including cell migration and embryogenesis (Johansson, S. et al. (1997) Front. Biosci. 2:D126-D146).

Lectins comprise a ubiquitous family of extracellular glycoproteins which bind cell surface carbohydrates specifically and reversibly, resulting in the agglutination of cells (reviewed in Drickamer, K. and Taylor, M. E. (1993) Annu. Rev. Cell Biol. 9:237-264). This function is particularly important for activation of the immune response. Lectins mediate the agglutination and mitogenic stimulation of lymphocytes at sites of inflammation (Lasky, L. A. (1991) J. Cell. Biochem. 45:139-146; Paietta, E. et al. (1989) J. Immunol. 143:2850-2857).

Lectins are further classified into subfamilies based on carbohydrate-binding specificity and other criteria. The galectin subfamily, in particular, includes lectins that bind β-galactoside carbohydrate moieties in a thiol-dependent manner (reviewed in Hadari, Y. R. et al. (1998) J. Biol. Chem. 270:3447-3453). Galectins are widely expressed and developmentally regulated. Galectins contain a characteristic carbohydrate recognition domain (CRD). The CRD comprises about 140 amino acids and contains several stretches of about 1 - 10 amino acids which are highly conserved among all galectins. A particular 6-amino acid motif within the CRD contains conserved tryptophan and arginine residues which are critical for carbohydrate binding. The CRD of some galectins also contains cysteine residues which may be important for disulfide bond formation. Secondary structure predictions indicate that the CRD forms several β-sheets.

Galectins play a number of roles in diseases and conditions associated with cell-cell and cell-matrix interactions. For example, certain galectins associate with sites of inflammation and bind to cell surface immunoglobulin E molecules. In addition, galectins may play an important role in cancer metastasis. Galectin overexpression is correlated with the metastatic potential of cancers in humans and mice. Moreover, anti-galectin antibodies inhibit processes associated with cell transformation, such as cell aggregation and anchorage-independent growth (see, for example, Su, Z.-Z. et al. (1996) Proc. Natl. Acad. Sci. USA 93:7252-7257).

Selectins, or LEC-CAMs, comprise a specialized lectin subfamily involved primarily in inflammation and leukocyte adhesion (Reviewed in Lasky, supra). Selectins mediate the recruitment of leukocytes from the circulation to sites of acute inflammation and are expressed on the surface of

vascular endothelial cells in response to cytokine signaling. Selectins bind to specific ligands on the leukocyte cell membrane and enable the leukocyte to adhere to and migrate along the endothelial surface. Binding of selectin to its ligand leads to polarized rearrangement of the actin cytoskeleton and stimulates signal transduction within the leukocyte (Brenner, B. et al. (1997) Biochem. Biophys. Res. Commun. 231:802-807; Hidari, K. I. et al. (1997) J. Biol. Chem. 272:28750-28756). Members of the selectin family possess three characteristic motifs: a lectin or carbohydrate recognition domain; an epidermal growth factor-like domain; and a variable number of short consensus repeats (scr or "sushi" repeats) which are also present in complement regulatory proteins.

Neural cell adhesion proteins (NCAPs) play roles in the establishment of neural networks during development and regeneration of the nervous system (Uyemura, K. et al. (1996) Essays Biochem. 31:37-48; Brummendorf, T., and Rathjen, F.G. (1996) Curr. Opin. Neurobiol. 6:584-593). NCAP participates in neuronal cell migration, cell adhesion, neurite outgrowth, axonal fasciculation, pathfinding, synaptic target-recognition, synaptic formation, myelination and regeneration. NCAPs are expressed on the surfaces of neurons associated with learning and memory. Mutations in genes encoding NCAPS are linked with neurological diseases, including hereditary neuropathy, Charcot-Marie-Tooth disease, Dejerine-Sottas disease, X-linked hydrocephalus, MASA syndrome (mental retardation, aphasia, shuffling gait and adducted thumbs), and spastic paraplegia type I. In some cases, expression of NCAP is not restricted to the nervous system. L1, for example, is expressed in melanoma cells and hematopoietic tumor cells where it is implicated in cell spreading and migration, and may play a role in tumor progression (Montgomery, A.M. et al. (1996) J. Cell Biol. 132:475-485).

NCAPs have at least one immunoglobulin constant or variable domain (Uyemura et al., supra). They are generally linked to the plasma membrane through a transmembrane domain and/or a glycosyl-phosphatidylinositol (GPI) anchor. The GPI linkage can be cleaved by GPI phospholipase C. Most NCAPs consist of an extracellular region made up of one or more immunoglobulin domains, a membrane spanning domain, and an intracellular region. Many NCAPs contain post-translational modifications including covalently attached oligosaccharide, glucuronic acid, and sulfate. NCAPs fall into three subgroups: simple-type, complex-type, and mixed-type. Simple-type NCAPs contain one or more variable or constant immunoglobulin domains, but lack other types of domains. Members of the simple-type subgroup include Schwann cell myelin protein (SMP), limbic system-associated membrane protein (LAMP), opiate-binding cell-adhesion molecule (OBCAM), and myelin-associated glycoprotein (MAG). The complex-type NCAPs contain fibronectin type III domains in addition to the immunoglobulin domains. The complex-type subgroup includes neural cell-adhesion molecule

(NCAM), axonin-1, F11, Bravo, and L1. Mixed-type NCAPs contain a combination of immunoglobulin domains and other motifs such as tyrosine kinase and epidermal growth factor-like domains. This subgroup includes Trk receptors of nerve growth factors such as nerve growth factor (NGF) and neurotropin 4 (NT4), Neu differentiation factors such as glial growth factor II (GGFII) and acetylcholine receptor-inducing factor (ARIA), and the semaphorin/collapsin family such as semaphorin B and collapsin.

Semaphorins are a large group of axonal guidance molecules consisting of at least 30 different members and are found in vertebrates, invertebrates, and even certain viruses. All semaphorins contain the sema domain which is approximately 500 amino acids in length. Neuropilin, a semaphorin receptor, has been shown to promote neurite outgrowth in vitro. The extracellular region of neuropilins consists of three different domains: CUB, discoidin, and MAM domains. The CUB and the MAM motifs of neuropilin have been proposed to have roles in protein-protein interactions and are suggested to be involved in the binding of semaphorins through the sema and the C-terminal domains (reviewed in Raper, J.A. (2000) Curr. Opin. Neurobiol. 10:88-94). Plexins are neuronal cell surface molecules that mediate cell adhesion via a homophilic binding mechanism in the presence of calcium ions. Plexins have been shown to be expressed in the receptors and neurons of particular sensory systems (Ohta, K. et al. (1995) Cell 14:1189-1199). There is evidence that suggests that some plexins function to control motor and CNS axon guidance in the developing nervous system. Plexins, which themselves contain complete semaphorin domains, may be both the ancestors of classical semaphorins and binding partners for semaphorins (Winberg, M.L. et al (1998) Cell 95:903-916).

An NCAP subfamily, the NCAP-LON subgroup, includes cell adhesion proteins expressed on distinct subpopulations of brain neurons. Members of the NCAP-LON subgroup possess three immunoglobulin domains and bind to cell membranes through GPI anchors. Kilon (a kindred of NCAP-LON), for example, is expressed in the brain cerebral cortex and hippocampus (Funatsu, N. et al. (1999) J. Biol. Chem. 274:8224-8230). Immunostaining localizes Kilon to the dendrites and soma of pyramidal neurons. Kilon has three C2 type immunoglobulin-like domains, six predicted glycosylation sites, and a GPI anchor. Expression of Kilon is developmentally regulated. It is expressed at higher levels in adult brain in comparison to embryonic and early postnatal brains. Confocal microscopy shows the presence of Kilon in dendrites of hypothalamic magnocellular neurons secreting neuropeptides, oxytocin or arginine vasopressin (Miyata, S. et al. (2000) J. Comp. Neurol. 424:74-85). Arginine vasopressin regulates body fluid homeostasis, extracellular osmolarity and intravascular volume. Oxytocin induces contractions of uterine smooth muscle during child birth and of myoepithelial cells in mammary glands during lactation. In magnocellular neurons, Kilon is proposed to

play roles in the reorganization of dendritic connections during neuropeptide secretion.

Cell adhesion proteins also include some members of the proline-rich proteins (PRPs). PRPs are defined by a high frequency of proline, ranging from 20-50% of the total amino acid content. Some PRPs have short domains which are rich in proline. These proline-rich regions are associated with protein-protein interactions. One family of PRPs are the proline-rich synapse-associated proteins (ProSAPs) which have been shown to bind to members of the postsynaptic density (PSD) protein family and subtypes of the somatostatin receptor (Yao, I. et al. (1999) J. Biol. Chem. 274: 27463-27466; Zitzer, H. et al. (1999) J. Biol. Chem. 274:32997-33001). Members of the ProSAP family contain six to seven ankyrin repeats at the N-terminus, followed by an SH3 domain, a PDZ domain, and seven proline-rich regions and a SAM domain at the C terminus. Several groups of ProSAPs are important structural constituents of synaptic structures in human brain (Zitzer et al., supra). Another member of the PRP family is the HLA-B-associated transcript 2 protein (BAT2) which is rich in proline and includes short tracts of polyproline, polyglycine, and charged amino acids. BAT2 also contains four RGD (Arg-Gly-Asp) motifs typical of integrins (Banerji, J. et al. (1990) Proc. Natl. Acad. Sci. USA 87:2374-2378).

Toposome is a cell-adhesion glycoprotein isolated from mesenchyme-blastula embryos. Toposome precursors including vitellogenin promote cell adhesion of dissociated blastula cells.

There are additional specific domains characteristic of cell adhesion proteins. One such domain is the MAM domain, a domain of about 170 amino acids found in the extracellular region of diverse proteins. These proteins all share a receptor-like architecture comprising a signal peptide, followed by a large N-terminal extracellular domain, a transmembrane region, and an intracellular domain (PROSITE document PDOC00604 MAM domain signature and profile). MAM domain proteins include zonadhesin, a sperm-specific membrane protein that binds to the zona pellucida of the egg; neuropilin, a cell adhesion molecule that functions during the formation of certain neuronal circuits, and Xenopus laevis thyroid hormone induced protein B, which contains four MAM domains and is involved in metamorphosis (Brown, D.D. et al. (1996) Proc. Natl. Acad. Sci. USA 93:1924-1929).

The WSC domain was originally found in the yeast WSC (cell-wall integrity and stress response component) proteins which act as sensors of environmental stress. The WSC domains are extracellular and are thought to possess a carbohydrate binding role (Ponting, C.P. et al. (1999) Curr. Biol. 9:S1-S2). A WSC domain has recently been identified in polycystin-1, a human plasma membrane protein. Mutations in polycystin-1 are the cause of the commonest form of autosomal dominant polycystic kidney disease (Ponting, C.P. et al. (1999) Curr. Biol. 9:R585-R588).

Leucine rich repeats (LRR) are short motifs found in numerous proteins from a wide range of species. LRR motifs are of variable length, most commonly 20-29 amino acids, and multiple repeats are typically present in tandem. LRR motifs are important for protein/protein interactions and cell adhesion, and LRR proteins are involved in cell/cell interactions, morphogenesis, and development (Kobe, B. and Deisenhofer, J. (1995) Curr. Opin. Struct. Biol. 5:409-416). The human ISLR (immunoglobulin superfamily containing leucine-rich repeat) protein contains a C2-type immunoglobulin domain as well as LRR motifs. The ISLR gene is linked to the critical region for Bardet-Biedl syndrome, a developmental disorder of which the most common feature is retinal dystrophy (Nagasawa, A. et al. (1999) Genomics 61:37-43).

The sterile alpha motif (SAM) domain is a conserved protein binding domain, approximately 70 amino acids in length, and is involved in the regulation of many developmental processes in eukaryotes. The SAM domain can potentially function as a protein interaction module through its ability to form homo- or hetero-oligomers with other SAM domains (Schultz, J. et al. (1997) Protein Sci. 6:249-253).

The discovery of new cell adhesion proteins, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of immune system disorders, neurological disorders, developmental disorders, and cell proliferative disorders, including cancer, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of cell adhesion proteins.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, cell adhesion proteins, referred to collectively as "CADHP" and individually as "CADHP-1," "CADHP-2," "CADHP-3," "CADHP-4," "CADHP-5," "CADHP-6," "CADHP-7," "CADHP-8," "CADHP-9," and "CADHP-10." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-10.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from

the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-10. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:11-20.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an

amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional CADHP, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional CADHP, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the

invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional CADHP, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the

compound.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble

polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"CADHP" refers to the amino acid sequences of substantially purified CADHP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of CADHP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other

compound or composition which modulates the activity of CADHP either by directly interacting with CADHP or by acting on components of the biological pathway in which CADHP participates.

An "allelic variant" is an alternative form of the gene encoding CADHP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding CADHP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as CADHP or a polypeptide with at least one functional characteristic of CADHP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding CADHP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding CADHP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent CADHP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of CADHP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of CADHP. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of CADHP either by directly interacting with CADHP or by acting on components of the biological pathway in which CADHP participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind CADHP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an in vitro evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH_2), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker. (See, e.g., Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13.)

The term "intramer" refers to an aptamer which is expressed in vivo. For example, a vaccinia

virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic CADHP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding CADHP or fragments of CADHP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide.

Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of CADHP or the polynucleotide encoding CADHP which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:11-20 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:11-20, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:11-20 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:11-20 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:11-20 and the region of SEQ ID NO:11-20 to which the fragment corresponds are routinely

determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-10 is encoded by a fragment of SEQ ID NO:11-20. A fragment of SEQ ID NO:1-10 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-10. For example, a fragment of SEQ ID NO:1-10 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-10. The precise length of a fragment of SEQ ID NO:1-10 and the region of SEQ ID NO:1-10 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other

polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default

parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized

after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of CADHP which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of CADHP which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of CADHP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of CADHP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an CADHP may involve lipidation, glycosylation,

phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of CADHP.

"Probe" refers to nucleic acid sequences encoding CADHP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer

selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing CADHP, nucleic acids encoding CADHP, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed

cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human cell adhesion proteins (CADHP), the polynucleotides encoding CADHP, and the use of these compositions for the diagnosis, treatment, or prevention of immune system disorders, neurological disorders, developmental disorders, and cell proliferative disorders, including cancer.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank homologs along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS

program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are cell adhesion proteins. For example, SEQ ID NO:1 is 46% identical, over 404 amino acid residues (I119-K522), to the murine, lectin C-type, Kupffer (hepatic sinusoid) cell receptor (GenBank ID g1669360, amino acid residues I106-K495), as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 2.1e-90. SEQ ID NO:1 is also 45% identical, over 414 amino acid residues (G109-K522), to the rat Kupffer cell receptor (GenBank ID g205051, amino acid residues G86-R495), as determined by BLAST analysis. The BLAST probability score is 7.2e-88. SEQ ID NO:1 also contains a lectin C-type domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from MOTIFS, BLIMPS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:1 is a cell adhesion protein.

In an alternative example, SEQ ID NO:2 is 61% identical, over 565 amino acid residues (A8-H569), to murine semaphorin VIa (GenBank ID g2623162, amino acid residues A5-H569), as determined by the Basic Local Alignment Search Tool (BLAST). The BLAST probability score is 6.6e-216. SEQ ID NO:2 also contains a semaphorin domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains.

In an alternative example, SEQ ID NO:3 is 44% identical, over 170 amino acid residues (L280-C449), to human integrin-binding protein Del-1 (GenBank ID g2865219, amino acid residues L307-C476), as determined by BLAST analysis. The BLAST probability score is 7.4e-32. SEQ ID NO:3 also contains a CUB domain (characteristic of developmentally-regulated proteins) and a f5/8 type C domain (characteristic of secreted proteins such as coagulation factors V and VIII, lactadherin, neuropilin-1, hemocytin, spondin, and discoidin I) as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains..

In an alternative example, SEQ ID NO:4 is 34% identical (over 329 amino acid residues) to human fibulin 1, isoform C (GenBank ID g31419), as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 6.2e-41, which indicates the

probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:4 is also 32% identical to human fibulin 2 (over 304 amino acid residues, GenBank ID g575233) and chicken fibulin 1, isoform C (over 329 amino acid residues, GenBank ID g2947316), with BLAST probability scores of 1.6e-36 and 1.4e-38, respectively. SEQ ID NO:4 also contains EGF-like domains, characteristic of fibulin polypeptides, as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:4 comprises EGF-like domains, consistent with the fibulin family of polypeptides.

In an alternative example, SEQ ID NO:5 is 72% identical (over 119 amino acid residues) to murine samaphorin (semaphorin) G (GenBank ID g1418942), as determined by BLAST analysis, with a probability score of 6.3e-39. SEQ ID NO:5 also contains a semaphorin domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains.

In an alternative example, SEQ ID NO:6 is 40% identical (over 597 amino acid residues) to murine MEG6, a high-molecular-weight protein with multiple EGF-like motifs (GenBank ID g3449294), as determined by BLAST analysis, with a probability score of 3.2e-149. SEQ ID NO:6 also contains EGF-like domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:6 comprises EGF-like and sushi domains, characteristic of members of the protein class.

In an alternative example, SEQ ID NO:8 is 98% identical over 1141 amino acid residues to human cadherin-23 (GenBank ID g11321508) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:8 also contains a cadherin domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:8 is a cadherin. SEQ ID NO:7, SEQ ID NO:9, and SEQ ID NO:10 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-10 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence

identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide sequences of the invention, and of fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:11-20 or that distinguish between SEQ ID NO:11-20 and related polynucleotide sequences.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotide sequences. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as FL_*XXXXXX*_N_1_N_2_*YYYYY*_N_3_N_4 represents a "stitched" sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYYY* is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FL*XXXXXX*_g*AAAAA*_g*BBBBB*_1_*N* is a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, g*AAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, g*BBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, *gBBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from

genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses CADHP variants. A preferred CADHP variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the CADHP amino acid sequence, and which contains at least one functional or structural characteristic of CADHP.

The invention also encompasses polynucleotides which encode CADHP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:11-20, which encodes CADHP. The polynucleotide sequences of SEQ ID NO:11-20, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding CADHP. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least

about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding CADHP. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:11-20 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:11-20. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of CADHP.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide sequence encoding CADHP. A splice variant may have portions which have significant sequence identity to the polynucleotide sequence encoding CADHP, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to the polynucleotide sequence encoding CADHP over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide sequence encoding CADHP. Any one of the splice variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of CADHP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding CADHP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring CADHP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode CADHP and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring CADHP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding CADHP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with

which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding CADHP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode CADHP and CADHP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding CADHP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:11-20 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding CADHP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.)

Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode CADHP may be cloned in recombinant DNA molecules that direct expression of CADHP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy

of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express CADHP.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter CADHP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of CADHP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding CADHP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, CADHP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of CADHP, or any part thereof, may be altered during direct synthesis and/or combined with

sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active CADHP, the nucleotide sequences encoding CADHP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding CADHP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding CADHP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding CADHP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding CADHP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding CADHP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors

(e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding CADHP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding CADHP can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding CADHP into the vector's multiple cloning site disrupts the *lac*Z gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of CADHP are needed, e.g. for the production of antibodies, vectors which direct high level expression of CADHP may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of CADHP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of CADHP. Transcription of sequences encoding CADHP may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding CADHP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses CADHP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of CADHP in cell lines is preferred. For example, sequences encoding CADHP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine

phosphoribosyltransferase genes, for use in *tk*⁻ and *aprˉ* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding CADHP is inserted within a marker gene sequence, transformed cells containing sequences encoding CADHP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding CADHP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding CADHP and that express CADHP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of CADHP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on CADHP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and

Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding CADHP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding CADHP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding CADHP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode CADHP may be designed to contain signal sequences which direct secretion of CADHP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding CADHP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric CADHP protein containing a heterologous moiety that can be recognized by a commercially available antibody may

facilitate the screening of peptide libraries for inhibitors of CADHP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the CADHP encoding sequence and the heterologous protein sequence, so that CADHP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled CADHP may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

CADHP of the present invention or fragments thereof may be used to screen for compounds that specifically bind to CADHP. At least one and up to a plurality of test compounds may be screened for specific binding to CADHP. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of CADHP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which CADHP binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express CADHP, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing CADHP or cell membrane fractions which contain CADHP are then contacted with a test compound and binding, stimulation, or inhibition of activity of either CADHP or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with CADHP, either in solution or affixed to a solid support, and detecting the binding of CADHP to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

CADHP of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of CADHP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for CADHP activity, wherein CADHP is combined with at least one test compound, and the activity of CADHP in the presence of a test compound is compared with the activity of CADHP in the absence of the test compound. A change in the activity of CADHP in the presence of the test compound is indicative of a compound that modulates the activity of CADHP. Alternatively, a test compound is combined with an in vitro or cell-free system comprising CADHP under conditions suitable for CADHP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of CADHP may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding CADHP or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous

strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding CADHP may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding CADHP can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding CADHP is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress CADHP, e.g., by secreting CADHP in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of CADHP and cell adhesion proteins. In addition, examples of tissues expressing CADHP can be found in Table 6. Therefore, CADHP appears to play a role in immune system disorders, neurological disorders, developmental disorders, and cell proliferative disorders, including cancer. In the treatment of disorders associated with increased CADHP expression or activity, it is desirable to decrease the expression or activity of CADHP. In the treatment of disorders associated with decreased CADHP expression or activity, it is desirable to increase the expression or activity of CADHP.

Therefore, in one embodiment, CADHP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CADHP. Examples of such disorders include, but are not limited to, an immune system disorder, such as acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, immunodeficiency associated with Cushing's disease, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia,

autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a developmental disorder, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and

cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; and a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, a vector capable of expressing CADHP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CADHP including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified CADHP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CADHP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of CADHP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CADHP including, but not limited to, those listed above.

In a further embodiment, an antagonist of CADHP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CADHP. Examples of such disorders include, but are not limited to, those immune system disorders, neurological disorders, developmental disorders, and cell proliferative disorders, including cancer, described above. In one aspect, an antibody which specifically binds CADHP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express CADHP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding CADHP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CADHP including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the

various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of CADHP may be produced using methods which are generally known in the art. In particular, purified CADHP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind CADHP. Antibodies to CADHP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with CADHP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to CADHP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of CADHP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to CADHP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single

chain antibodies may be adapted, using methods known in the art, to produce CADHP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for CADHP may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between CADHP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering CADHP epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for CADHP. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of CADHP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple CADHP epitopes, represents the average affinity, or avidity, of the antibodies for CADHP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular CADHP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the CADHP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of CADHP, preferably in active form, from the

antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of CADHP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding CADHP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding CADHP. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding CADHP. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding CADHP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-

linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in CADHP expression or regulation causes disease, the expression of CADHP from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in CADHP are treated by constructing mammalian expression vectors encoding CADHP and introducing these vectors by mechanical means into CADHP-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of CADHP include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). CADHP may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen));

the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding CADHP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to CADHP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding CADHP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding CADHP to cells which have one or more genetic abnormalities with respect

to the expression of CADHP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding CADHP to target cells which have one or more genetic abnormalities with respect to the expression of CADHP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing CADHP to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding CADHP to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA,

resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for CADHP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of CADHP-coding RNAs and the synthesis of high levels of CADHP in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of CADHP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding CADHP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary

oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding CADHP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding CADHP. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased CADHP expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding CADHP may be therapeutically useful, and in the treatment of disorders associated with decreased CADHP expression or activity, a compound which specifically promotes expression of the polynucleotide encoding CADHP may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound

based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding CADHP is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding CADHP are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding CADHP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of

CADHP, antibodies to CADHP, and mimetics, agonists, antagonists, or inhibitors of CADHP.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising CADHP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, CADHP or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example CADHP or fragments thereof, antibodies of CADHP, and agonists, antagonists or inhibitors of CADHP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the

dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind CADHP may be used for the diagnosis of disorders characterized by expression of CADHP, or in assays to monitor patients being treated with CADHP or agonists, antagonists, or inhibitors of CADHP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for CADHP include methods which utilize the antibody and a label to detect CADHP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring CADHP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of CADHP expression. Normal or standard values for CADHP expression are established by combining body fluids or cell

extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to CADHP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of CADHP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding CADHP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of CADHP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of CADHP, and to monitor regulation of CADHP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding CADHP or closely related molecules may be used to identify nucleic acid sequences which encode CADHP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding CADHP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the CADHP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:11-20 or from genomic sequences including promoters, enhancers, and introns of the CADHP gene.

Means for producing specific hybridization probes for DNAs encoding CADHP include the cloning of polynucleotide sequences encoding CADHP or CADHP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding CADHP may be used for the diagnosis of disorders associated with expression of CADHP. Examples of such disorders include, but are not limited to, an immune system disorder, such as acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome

(thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, immunodeficiency associated with Cushing's disease, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a developmental disorder, such as renal tubular acidosis, anemia,

Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; and a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. The polynucleotide sequences encoding CADHP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered CADHP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding CADHP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding CADHP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding CADHP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of CADHP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding CADHP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified

polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding CADHP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding CADHP, or a fragment of a polynucleotide complementary to the polynucleotide encoding CADHP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding CADHP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding CADHP are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP

(isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of CADHP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, CADHP, fragments of CADHP, or antibodies specific for CADHP may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No.

5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be

quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for CADHP to quantify the levels of CADHP expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and

should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding CADHP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific

region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding CADHP on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, CADHP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between CADHP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are

synthesized on a solid substrate. The test compounds are reacted with CADHP, or fragments thereof, and washed. Bound CADHP is then detected by methods well known in the art. Purified CADHP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding CADHP specifically compete with a test compound for binding CADHP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with CADHP.

In additional embodiments, the nucleotide sequences which encode CADHP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/256,542, U.S. Ser. No. 60/259,604, and U.S. Ser. No. 60/260,101, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 3. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles

(QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from Homo sapiens, Rattus norvegicus, Mus musculus, Caenorhabditis elegans, Saccharomyces cerevisiae, Schizosaccharomyces pombe, and Candida albicans (Incyte Genomics, Palo Alto CA); and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide

sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:11-20. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative cell adhesion proteins were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode cell adhesion proteins, the encoded polypeptides were analyzed by querying against PFAM models for cell adhesion proteins. Potential cell adhesion proteins were also

identified by homology to Incyte cDNA sequences that had been annotated as cell adhesion proteins. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of CADHP Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:11-20 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:11-20 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum}\{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding CADHP are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system;

pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding CADHP. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of CADHP Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II

(Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:11-20 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase

(DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described

in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/µl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/µl RNase inhibitor, 500 µM dATP, 500 µM dGTP, 500 µM dTTP, 40 µM dCTP, 40 µM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37° C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85° C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 µl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 µg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 µl of the array element DNA, at an average concentration of 100 ng/µl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60° C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 µl of sample mixture consisting of 0.2 µg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65° C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 µl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60° C. The arrays are washed for 10 min at 45° C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45° C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from

different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XI. Complementary Polynucleotides

Sequences complementary to the CADHP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring CADHP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of CADHP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the CADHP-encoding transcript.

XII. Expression of CADHP

Expression and purification of CADHP is achieved using bacterial or virus-based expression systems. For expression of CADHP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express CADHP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of CADHP in eukaryotic cells is achieved by infecting

insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding CADHP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, CADHP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from CADHP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified CADHP obtained by these methods can be used directly in the assays shown in Examples XVI and XVII, where applicable.

XIII. Functional Assays

CADHP function is assessed by expressing the sequences encoding CADHP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an

automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of CADHP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding CADHP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding CADHP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIV. Production of CADHP Specific Antibodies

CADHP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the CADHP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-CADHP activity by, for example, binding the peptide or CADHP to a substrate,

blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XV. Purification of Naturally Occurring CADHP Using Specific Antibodies

Naturally occurring or recombinant CADHP is substantially purified by immunoaffinity chromatography using antibodies specific for CADHP. An immunoaffinity column is constructed by covalently coupling anti-CADHP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing CADHP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of CADHP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/CADHP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and CADHP is collected.

XVI. Identification of Molecules Which Interact with CADHP

CADHP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled CADHP, washed, and any wells with labeled CADHP complex are assayed. Data obtained using different concentrations of CADHP are used to calculate values for the number, affinity, and association of CADHP with the candidate molecules.

Alternatively, molecules interacting with CADHP are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

CADHP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVII. Demonstration of CADHP Activity

An assay for CADHP activity measures the expression of CADHP on the cell surface. cDNA encoding CADHP is transfected into a non-leukocytic cell line. Cell surface proteins are labeled with biotin (de la Fuente, M.A. et al. (1997) Blood 90:2398-2405). Immunoprecipitations are performed using CADHP-specific antibodies, and immunoprecipitated samples are analyzed using SDS-PAGE and immunoblotting techniques. The ratio of labeled immunoprecipitant to unlabeled

immunoprecipitant is proportional to the amount of CADHP expressed on the cell surface.

Alternatively, an assay for CADHP activity measures the amount of cell aggregation induced by overexpression of CADHP. In this assay, cultured cells such as NIH3T3 are transfected with cDNA encoding CADHP contained within a suitable mammalian expression vector under control of a strong promoter. Cotransfection with cDNA encoding a fluorescent marker protein, such as Green Fluorescent Protein (CLONTECH), is useful for identifying stable transfectants. The amount of cell agglutination, or clumping, associated with transfected cells is compared with that associated with untransfected cells. The amount of cell agglutination is a direct measure of CADHP activity.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
4350981	1	4350981CD1	11	4350981CB1
7596315	2	7596315CD1	12	7596315CB1
71234712	3	71234712CD1	13	71234712CB1
079370	4	079370CD1	14	079370CB1
2496174	5	2496174CD1	15	2496174CB1
4097936	6	4097936CD1	16	4097936CB1
2523646	7	2523646CD1	17	2523646CB1
4099073	8	4099073CD1	18	4099073CB1
7156379	9	7156379CD1	19	7156379CB1
7473626	10	7473626CD1	20	7473626CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
1	4350981CD1	g1669360	2.1e-90	Kupffer cell receptor [Mus musculus].
2	7596315CD1	g2623162	6.6e-216	Semaphorin VIa [Mus musculus]. Zhou, L. et al. (1997) Cloning and expression of a novel murine semaphorin with structural similarity to insect semaphorin I. Mol. Cell. Neurosci. 9:26-41.
3	71234712CD1	g2865219	7.4e-32	Integrin binding protein Del-1 [Homo sapiens]. Hidai, C. et al. (1998) Cloning and characterization of developmental endothelial locus-1: An embryonic endothelial cell protein that binds the alphavbeta3 integrin receptor. Genes Dev. 12:21-33.
4	079370CD1	g31419	6.2e-41	Fibulin-1 C [Homo sapiens]. Argraves, W.S. et al. (1990) Fibulin is an extracellular matrix and plasma glycoprotein with repeated domain structure. J. Cell Biol. 111:3155-3164.
5	2496174CD1	g1418942	6.3e-39	Samaphorin G (semaphorin) [Mus musculus]. Adams, R.H. et al. (1996) A novel class of murine semaphorins with homology to thrombospondin is differentially expressed during early embryogenesis. Mech. Dev. 57:33-45.
6	4097936CD1	g3449294	3.2e-149	MEGF6 [Rattus norvegicus]. Nakayama, M., et al. (1998) Identification of high-molecular-weight proteins with multiple EGF-like motifs by motif-trap screening. Genomics 51:27-34.
7	2523646CD1	g11321508	4.5e-117	[Homo sapiens] (AY010111) cadherin-23 Bork, J.M. et al. (2001) Am. J. Hum. Genet. 68:26-37.
8	4099073CD1	g11321508	0.0	[Homo sapiens] (AY010111) cadherin-23 Bork, J.M. et al. (2001) Am. J. Hum. Genet. 68:26-37.
9	7156379CD1	g1655432	0.0	[Mus musculus] plexin 2 Kameyama, T. et al. (1996) Biochem. Biophys. Res. Commun. 226:396-402.

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
10	7473626CD1	g693910	1.9e-38	[Mus musculus] seizure-related gene product 6 precursor Shimizu-Nishikawa, K. et al. (1995) Brain Res. Mol. Brain. Res. 28:201-210.

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	4350981CD1	523	T76, T115, S141, T146, S173, T178, S216, S233, S278, T283, T315, S334, T362, T423, S455, S459, S479, T490, T500, S507	N83, N177, N200, N214, N282, N355, N369	Signal cleavage: M1-H61	SPSCAN
					Signal Peptide: M45-F64, M27-L56, M45-E67, M27-V58	HMMER
					Signal cleavage: M1-H63	PROFILESCAN
					Lectin C-type domain: Q465-K522	HMMER-PFAM
					Transmembrane domain: P32-D60, N-terminus is non-cytoplasmic	TMAP
					PROTEIN KUPFFER CELL RECEPTOR TRANSMEMBRANE GLYCOPROTEIN LECTIN SIGNALANCHOR ENDOCYTOSIS GOLGIN160 PD031152: K123-E464, M1-Q365, R66-K368	BLAST-PRODOM
					C-TYPE LECTIN DM00035|P10716|405-536: Q432-K522	BLAST-DOMO
					Type II antifreeze proteins (contain C-type lectin domain): PR00356: L450-C467, V468-F485, W496-D512	BLIMPS-PRINTS
					Cell attachment sequence (RGD): R395-D397	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
2	7596315CD1	1017	S22, S49, Y70, T97, S151, T187, S201, T210, S266, S299, S332, T381, S459, S513, S520, T576, S650, S678, T687, S688, S734, S736, S745, T749, S808, S809, T822, S858, T886, T900, S913, S991, S1008	N51, N283, N435, N461, N776, N782, N911, N978	Signal Peptide: M1-A20	HMMER
					Semaphorin domain: F59-C477	HMMER-PFAM
					Transmembrane domains: F4-V21, M602-Y630, N-terminus is cytoplasmic	TMAP
					SEMAPHORIN PROTEIN PRECURSOR RECEPTOR KINASE SIGNAL TYROSINE TYROSINEPROTEIN FAMILY HEPATOCYTE PD001844: K249-K476, S161-I300, D67-K182, R988-Y1017	BLAST-PRODOM
					SEMAPHORIN; FASCICLIN; COLLAPSIN; II; DM01606|JH0798|1-640: M1-V603	BLAST-DOMO
3	71234712CD1	561	T97, S140, S176, S185, T194, T286, S312, T320, S325, T362, T370, S497	N41, N95, N155, N272, N474, N516, N522	Signal cleavage: M1-A66	SPSCAN
					Signal Peptide: M47-D63, M47-A64, M47-A66, M47-Q68, N41-A66, S40-A66, M47-A66, M47-Q67, M47-Q68, M47-D70, S42-Q68, C39-Q68	HMMER
					CUB domain: C72-Y184	HMMER-PFAM
					F5/8 type C domain: T295-L446	HMMER-PFAM
					Transmembrane domains: N41-G65 T524-W552	TMAP
					GLYCOPROTEIN PRECURSOR SIGNAL FACTOR REPEAT PROTEIN NEUROPILIN CELL DOMAIN COAGULATION PD000875: L296-L446	BLAST-PRODOM
					DISCOIDIN I N-TERMINAL DM00516|A42580|2085-2210: P328-C449	BLAST-DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4	079370CD1	439	S5, T86, T150, S250, T261, S267, S370, S389, S435	N124, N307	Signal cleavage: M1-A23	SPSCAN
					Signal Peptide: M1-P21, M1-A23, M1-S29, M1-Q25, M1-C27	HMMER
					EGF-like domain: C140-C171, C228-C268, C274-C318	HMMER-PFAM
					Sushi domain (SCR repeat): C81-C134	HMMER-PFAM
					Type II EGF-like signature: PR00010: G136-N147, G148-V155, G249-Y259, D264-D270	BLIMPS-PRINTS
					Sushi domain proteins: PF00084: D100-F111, G125-C134	BLIMPS-PFAM
					EGF-LIKE DOMAIN GLYCOPROTEIN PRECURSOR SIGNAL EXTRACELLULAR MATRIX PLASMA CALCIUM-BINDING REPEAT: PD008104: Q317-V434 (p = 3.4e-09)	BLAST-PRODOM
					EGF DM00003|P18168|918-958: I137-Q172, V271-G304, V225-G258 (p = 1.7e-10)	BLAST-DOMO
					Aspartic acid and asparagine hydroxylation site (characteristic of EGF-related proteins): C244-C255	MOTIFS
					EGF-like domain signature 1: C160-C171 EGF-like domain signature 2: C253-C268	MOTIFS
					Calcium-binding EGF-like domain pattern signature: D224-C253, D270-C296	MOTIFS
5	2496174CD1	160	S25, S153, T156	N59, N95	Signal cleavage: M1-S26	SPSCAN
					Signal Peptide: M1-S25, M1-D28, M1-Q27, M1-P33	HMMER
					Semaphorin (sema) domain: F68-E101	HMMER-PFAM
					Transmembrane domain: A4-S21, N-terminus is non-cytosolic	TMAP

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					SEMAPHORIN G PRECURSOR SAMAPHORIN G SIGNAL: PD107004: Q27-D67	BLAST-PRODOM
6	4097936CD1	994	S30, T38, S154, T346, Y355, S448, S460, S535, S566, T581, T644, S649, S753, S840, S841, S865, T884, S899, S921, S966	N152, N153, N271, N392, N446 N476, N491, N575, N626, N634, N660, N817	Signal cleavage: M1-T20	SPSCAN
					Signal Peptide: M1-R16, M1-A18, M1-G19, M1-T20, M1-N22, M1-S24, M1-D25, M1-T28, M1-C29	HMMER
					EGF-like domain: C101-C131, C144-C174, C187-C216, C229-C259, C272-C302, C315-C345, C365-C391, C404-C434, C447-C477, C490-C520, C533-C563, C576-C606, C619-C648, C661-C691	HMMER-PFAM
					Transmenbrane domains: A708-R736, N-terminus is cytosolic	TMAP
					Type III EGF-like signature: PR00011: C284-C302, C284-C302, S133-G161, C284-C302	BLIMPS-PRINTS
					Sushi domain proteins: PF00084 C248-P252, G603-R614 (p = 0.0049)	BLIMPS-PFAM
					SUSHI REPEAT DM04887|P16581|1-609: G138-C576	BLAST-DOMO
					Cell attachment sequence (RGD): R127-D129	MOTIFS
					EGF-like domain signature 1: C120-C131, C205-C216, C248-C259, C291-C302, C334-C345, C380-C391, C423-C434, C509-C520, C552-C563, C595-C606, C637-C648, C680-C691 EGF-like domain signature 2: C120-C131, C248-C259, C291-C302, C334-C345, C380-C391, C423-C434, C466-C477, C509-C520, C552-C563, C595-C606, C637-C648	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
7	2523646CD1	987	S82 S284 S336 S368 S508 S513 S579 S632 S759 S826 S929 T57 T108 T154 T159 T213 T224 T385 T391 T444 T486 T596 T638 T667 T677 T691 T730 T849 T851 T946	N55 N106 N249 N293 N334 N366 N372 N552 N594 N665 N710 N727 N841 N927 N944	Cadherin domain: Y365-L452, F795-C886, Y466-F562, L37-T127, Y576-L669, Y683-N779, Y141-L237, Y253-I351	HMMER_PFAM
					Transmembrane domain: S254-V273, Y576-N594 N-terminus is cytosolic	TMAP
					Cadherins extracellular repeat proteins domain proteins BL00232: Y343-G390, T659-P676	BLIMPS_BLOCKS
					Cadherins extracellular repeated domain signature: L536-V590, G216-L267, T430-L480, I756-I809, I643-V697	PROFILESCAN
					Cadherin signature PR00205: V375-G390, T659-P676, V700-V714	BLIMPS_PRINTS
					INSECTICIDAL TOXIN RECEPTOR BTR1 PRECURSOR RECEPTOR GLYCOPROTEIN TRANSMEMBRANE SIGNAL REPEAT CELL ADHESION PD134331: P74-Q462, A419-I883, T432-P794	BLAST_PRODOM
					CADHERIN REPEAT DM00030|P33450|2417-2519: E394-D489, P491-D599 DM00030|P33450|1952-2055: G390-S490 DM00030|I48277|191-296: Y604-L707 DM00030|P33450|3576-3680: F619-L707	BLAST_DOMO
					Cadherins extracellular repeated domain signature I236-P246 I449-P459 I559-P569 I666-P676 V778-P788	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
8	4099073CD1	1619	S89 S150 S163 S244 S295 S439 S470 S482 S559 S706 S740 S761 S786 S817 S1041 S1107 S1355 S1530 S1564 S1596 T50 T80 T227 T392 T394 T447 T497 T520 T634 T671 T725 T727 T753 T834 T867 T938 T1036 T1128 T1147 T1318 T1420 T1515 T1581 T1586 T1612 Y689 Y1487	N19 N35 N186 N225 N257 N270 N381 N429 N468 N495 N563 N657 N669 N677 N916 N1049 N1108 N1177 N1196 N1241 N1281 N1455	Cadherin domain: G118-L210, F814-I902, Y7-N103, Y707-L800, Q374-D465, M224-I315, T918-E1013, Q1033-T1128, Y601-V693, Q479-L584	HMMER_PFAM
					Transmembrane domains: A412-G428, I1175-L1194, D1211-R1235, D1360-W1388 N-terminus is non-cytosolic	TMAP
					Cadherins extracellular repeated domain signature: T443-I493, T671-I721, F184-V238, S878-L928, L292-L346	PROFILESCAN
					INSECTICIDAL TOXIN RECEPTOR BTR1 PRECURSOR RECEPTOR GLYCOPROTEIN TRANSMEMBRANE SIGNAL REPEAT CELL ADHESION PD131770: F702-I1264, V460-T727 PD134331: G385-S815	BLAST_PRODOM
					SIMILARITY TO MULTIPLE CADHERIN-TYPE REPEATS CELL ADHESION GLYCOPROTEIN TRANSMEMBRANE CALCIUM-BINDING REPEAT PD131836: P700-P909	BLAST_PRODOM
					CADHERIN REPEAT DM00030|P33450|1079-1181: G142-D247, G249-D359, G732-D837 DM00030|P33450|417-522: G142-S248, G626-S731 DM00030|P33450|1952-2055: G652-S731, G142-S248 DM00030|P33450|3259-3362: G142-S248, G249-T344	BLAST_DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Cadherins extracellular repeated domain signature V100-P110 I207-P217 I315-P325 V581-P591 L690-P700 V899-P909	MOTIFS
9	7156379CD1	1894	S34 S177 S200 S201 S208 S292 S303 S349 S452 S498 S515 S599 S664 S689 S1010 S1264 S1305 S1374 S1379 S1396 S1432 S1484 S1543 S1619 S1633 S1765 S1783 S1795 S1806 S1825 T42 T86 T103 T187 T205 T270 T277 T399 T537 T594 T721 T833 T877 T955 T1009 T1063 T1112 T1193 T1270 T1321 T1571 T1572 T1737 T1777 Y495 Y1343 Y1815	N7 N164 N442 N567 N592 N655 N756 N764 N1007 N1090 N1132 N1135 N1180 N1609 N1610	Plexin repeat: S509-V559, N655-P702, K803-T856	HMMER_PFAM
					Sema domain: F51-D490	HMMER_PFAM
					IPT/TIG domain: P858-M952, L954-V1037, P1040-Y1139	HMMER_PFAM
					Transmembrane domain: S1232-Y1260, K1353-F1368 N-terminus is non-cytosolic	TMAP
					PLEXIN PRECURSOR SIGNAL PROTEIN KIAA0407 TRANSMEMBRANE SEX RECEPTOR GLYCOPROTEIN VESPR PD008852: I1300-S1894, A1255-S1670	BLAST_PRODOM
					RECEPTOR KINASE TYROSINE PROTEIN PRECURSOR TYROSINEPROTEIN SIGNAL HEPATOCYTE GROWTH ATPBINDING PD003981: E912-N1205, C855-S945	BLAST_PRODOM
					PLEXIN PROTEIN PRECURSOR SIGNAL KIAA0407 K04B12.1 TRANSMEMBRANE SEX RECEPTOR GLYCOPROTEIN PD010132: P565-H837	BLAST_PRODOM
					PLEXIN PRECURSOR SIGNAL TRANSMEMBRANE PROTEIN SEX RECEPTOR GLYCOPROTEIN PD003973: R370-H491	BLAST_PRODOM
					do KINASE; TYROSINE; HEPATOCYTE; ATP; DM03653|P08581|14-526: H62-C516 DM03653|A48196|13-528: I63-C516	BLAST_DOMO
					do KINASE; TYROSINE; ATP; GROWTH; DM01368|P51805|796-899: C814-E920 DM02937|P51805|991-1085: V1019-L1103	BLAST_DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
10	7473626CD1	326	S41 S46 S90 S213 T4 T22 T82 T101 T176 T246	N63 N194 N199 N232	CUB domain: C19-Y124, C191-F295	HMMER_PFAM
					Sushi domain (SCR repeat): C132-C187	HMMER_PFAM
					Transmembrane domains: C45-I73, L279-W305 N-terminus is cytosolic	TMAP
					GLYCOPROTEIN DOMAIN EGFLIKE PROTEIN PRECURSOR SIGNAL RECEPTOR INTRINSIC FACTORB12 REPEAT PD000165: C19-Y124	BLAST_PRODOM
					C1R/C1S REPEAT DM00162|I49540|592-708: C15-Y124 DM00162|P98069|418-529: A17-Y124, C191-A292 DM00162|A57190|826-947: W8-Y124 DM00162|P98066|131-247: Q18-A126	BLAST_DOMO

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
11/ 4350981CB1/ 2052	1-1566, 119-182, 179-641, 179-681, 457-1023, 966-1191, 966-1235, 966-1376, 966-1401, 966-1412, 966-1413, 966-1414, 966-1433, 966-1442, 966-1444, 966-1452, 966-1470, 966-1483, 966-1484, 966-1485, 966-1489, 966-1494, 966-1507, 966-1508, 966-1524, 966-1537, 966-1639, 977-1395, 1004-1575, 1015-1680, 1038-1662, 1048-1588, 1054-1587, 1076-1552, 1092-1605, 1124-1581, 1153-1639, 1173-1492, 1199-1659, 1218-1752, 1218-1800, 1239-1537, 1250-1552, 1257-1713, 1267-1653, 1301-1934, 1304-1663, 1308-1639, 1334-2033, 1342-2052, 1363-1944, 1379-2036, 1398-2012, 1400-1585, 1425-2034, 1477-1921, 1501-2052, 1548-2052, 1556-2052, 1562-2047, 1580-2052, 1583-2049
12/ 7596315CB1/ 4234	1-413, 28-318, 28-464, 33-662, 406-1063, 407-582, 407-660, 407-846, 408-501, 408-660, 410-910, 411-910, 568-865, 587-667, 607-910, 682-745, 744-809, 761-1295, 775-809, 801-1419, 809-893, 809-1001, 915-1357, 973-1620, 982-1620, 1082-1372, 1086-1699, 1149-1604, 1208-1620, 1259-1829, 1559-2101, 1891-2224, 1891-2295, 1891-2320, 1897-2109, 1943-2033, 2001-2729, 2025-2109, 2032-2109, 2107-2674, 2107-2723, 2113-2735, 2120-2170, 2159-2228, 2191-2735, 2227-2317, 2227-2481, 2227-2563, 2228-2802, 2230-2818, 2302-2668, 2323-2903, 2504-3117, 2516-2842, 2516-2990, 2619-3144, 2711-3144, 2730-2964, 2730-3261, 2848-3131, 2913-3367, 2986-3527, 2986-3642, 3189-3858, 3259-3837, 3338-3864, 3363-3858, 3377-4005, 3410-3756, 3536-4114, 3540-4123, 3542-4129, 3568-3829, 3624-4002, 3717-4223, 3729-3905, 3729-4234, 3753-4022
13/ 71234712CB1/ 2200	1-409, 1-434, 10-658, 15-559, 30-409, 40-548, 41-375, 92-409, 403-1007, 405-815, 410-1041, 411-880, 412-895, 440-830, 478-1047, 480-845, 480-853, 480-897, 480-905, 480-912, 486-792, 492-1044, 495-778, 495-923, 495-1027, 497-849, 497-1077, 600-1097, 722-1278, 771-1036, 795-1371, 797-1415, 798-1350, 823-1370, 826-1407, 832-1199, 840-1401, 843-1425, 854-1307, 864-1260, 872-1385, 878-1422, 881-1557, 934-1469, 966-1560, 969-1602, 973-1413, 974-1576, 990-1371, 994-1251, 994-1602, 1009-1534, 1015-1274, 1020-1189, 1051-1278, 1121-1583, 1134-1384, 1141-1781, 1150-1437, 1151-1431, 1162-1343, 1185-1784, 1200-1471, 1247-1662, 1252-1730, 1255-1905, 1269-1846, 1305-1773, 1347-1959, 1388-1998, 1390-1554, 1415-2016, 1419-1964, 1422-2061, 1439-2075, 1442-2055, 1464-1749, 1489-1763, 1508-1810, 1555-1795, 1567-1816, 1567-1819, 1571-2200, 1586-1833, 1595-1881, 1638-1779, 1662-2115, 1827-2084, 1875-2095

Table 4 (cont.)

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
14/ 079370CB1/ 1647	1-160, 74-729, 193-515, 212-303, 492-1405, 618-829, 728-1051, 728-1261, 768-1062, 786-1024, 877-1450, 877-1508, 952-1309, 1033-1504, 1038-1647, 1137-1427, 1137-1642, 1305-1520, 1375-1641, 1416-1647
15/ 2496174CB1/ 4456	1-305, 166-827, 272-2407, 506-847, 611-851, 883-1016, 883-1184, 1471-1926, 1545-2162, 1799-2311, 1799-2384, 1880-2435, 2332-2636, 2332-2813, 2428-2539, 2539-2691, 2539-4443, 2585-2831, 2586-2775, 2586-3123, 3112-3438, 3121-3360, 3121-3678, 3153-3415, 3350-3992, 3387-3585, 3387-3927, 3429-4008, 3450-3978, 3479-4149, 3519-4039, 3528-3952, 3528-3963, 3528-3964, 3528-4051, 3528-4126, 3540-4093, 3551-4190, 3607-4244, 3613-4320, 3621-4289, 3661-4386, 3685-4335, 3691-4283, 3695-4394, 3800-4408, 3803-3978, 3808-4367, 3819-4423, 3839-4331, 3865-4454, 3868-4421, 3880-4103, 3903-4184, 3904-4404, 3904-4406, 3922-4344, 3925-4452, 3927-4432, 3937-4456, 3941-4453, 3944-4203
16/ 4097936CB1/ 3574	1-22, 1-24, 1-738, 233-841, 243-3227, 287-965, 783-1548, 814-1562, 1099-1245, 1099-1730, 1114-1367, 1148-1552, 1150-1552, 1157-1418, 1199-1562, 1356-1962, 1358-1532, 1587-1859, 1692-1925, 1894-2174, 2194-2365, 2194-2556, 2194-2677, 2194-2694, 2248-2654, 2296-2585, 2323-2592, 2350-2609, 2350-2817, 2375-2977, 2467-3017, 2467-3183, 2469-3123, 2537-3001, 2625-2999, 2636-3195, 2645-3510, 2683-2940, 2683-2948, 2683-3011, 2726-3510, 2733-3510, 2743-3510, 2802-2927, 2802-2936, 2806-3574, 2882-3112, 2901-3143, 2906-3255, 2926-3574, 2930-3532, 2930-3545, 2930-3574, 2992-3251, 3040-3305, 3123-3343
17/ 2523646CB1/ 3562	1-537, 340-745, 665-1272, 745-995, 745-1300, 777-989, 777-1316, 922-1201, 992-1192, 1044-1639, 1432-1773, 1508-1938, 1718-2091, 1718-2159, 1718-2187, 1718-2219, 1718-2222, 1718-2285, 1718-2318, 1718-2399, 1770-2310, 1780-2273, 1789-2419, 1814-2421, 1820-1997, 1820-2043, 1820-2157, 1820-2179, 1820-2231, 1820-2276, 1820-2305, 1820-2379, 1827-2026, 1975-2262, 1975-2585, 2032-2538, 2047-2184, 2101-2723, 2136-2669, 2159-2399, 2159-2561, 2195-2434, 2198-2633, 2201-2725, 2243-2787, 2284-2571, 2298-2542, 2325-2943, 2343-2832, 2423-2915, 2450-2869, 2451-2943, 2468-2607, 2542-2943, 2558-2943, 2643-2943, 2700-2943, 2738-2943, 2766-2943, 2901-2941, 2940-3562, 2943-2990, 2943-3064, 2943-3163, 2943-3313, 2943-3346, 2943-3494, 2943-3559, 2944-2999, 2955-3545, 2966-3270, 2975-3508, 3008-3292, 3008-3523, 3008-3530, 3114-3416, 3216-3412, 3216-3420, 3224-3482

Table 4 (cont.)

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
18/ 4099073CB1/ 6197	1-939, 43-138, 44-138, 65-138, 75-138, 81-138, 102-138, 111-172, 177-586, 192-301, 209-946, 250-946, 294-932, 298-956, 303-939, 327-939, 329-944, 355-944, 365-944, 368-944, 369-931, 369-956, 371-944, 375-939, 375-944, 394-944, 405-939, 420-944, 421-944, 422-945, 428-934, 429-944, 448-944, 452-957, 461-944, 469-944, 472-956, 475-944, 480-944, 493-944, 513-944, 518-1017, 540-934, 550-944, 551-944, 568-944, 575-957, 577-957, 580-944, 628-939, 633-944, 637-957, 648-940, 649-963, 660-939, 664-934, 666-939, 667-942, 670-939, 670-943, 675-933, 679-934, 689-943, 690-946, 694-929, 694-936, 697-937, 697-938, 697-940, 698-933, 700-939, 702-934, 704-946, 704-949, 704-957, 708-932, 710-942, 711-945, 717-1014, 721-946, 722-939, 722-944, 722-946, 729-944, 730-948, 735-949, 736-934, 757-934, 787-940, 787-950, 788-939, 791-6008, 799-996, 1143-1803, 1237-1778, 2081-2197, 2081-2205, 2081-2698, 2191-2655, 2294-2912, 2385-2760, 2440-2739, 2575-3050, 2608-3124, 2613-3151, 2648-3385, 2651-3246, 2660-3242, 2756-3144, 2756-3383, 2796-3633, 2815-3484, 2898-3545, 2989-3641, 2993-3379, 3012-3657, 3014-3333, 3032-3699, 3065-3957, 3188-3851, 3189-3964, 3238-3533, 3334-4052, 3386-4163, 3414-4027, 3512-4165, 3532-4021, 3548-4203, 3567-3921, 3628-4454, 3682-4176, 3813-4480, 3850-4492, 3905-4437, 3989-4689, 4012-4132, 4012-4614, 4012-4666, 4017-4334, 4075-4734, 4107-4670, 4149-4816, 4159-4746, 4176-4423, 4186-4792, 4205-4611, 4256-4539, 4285-4726, 4330-4778, 4354-4477, 4438-5047, 4448-4917, 4488-5143, 4489-4609, 4572-5174, 4622-5352, 4646-5202, 4664-5229, 4693-5150, 4723-5476, 4760-5209, 4768-5257, 4770-5330, 4779-5471, 4783-5617, 4785-5372, 4792-5378, 4798-5031, 4798-5204, 4798-5239, 4799-5432, 4802-5394, 4820-5364, 4825-5449, 4857-5320, 4870-5465, 4883-5430, 4931-5010, 4937-5491, 4954-5152, 4954-5396, 4954-5424, 4954-5446, 4954-5505, 4984-5469, 4985-5587, 4996-5404, 5018-5572, 5026-5095, 5027-5279, 5028-5163, 5076-5583, 5085-5577, 5149-5724, 5163-5681, 5163-5851, 5210-5469, 5210-5732, 5216-5777, 5236-5681, 5250-5893, 5273-5394, 5280-5547, 5280-5672, 5298-5681, 5305-5984, 5320-5620, 5320-5800, 5320-5889, 5326-5979, 5340-5681, 5353-5681, 5356-5681, 5365-5956, 5367-5633, 5368-6000, 5371-5947, 5383-5924, 5398-5891, 5410-5681, 5428-5559, 5500-5681, 5513-6078, 5540-5617, 5547-5968, 5549-5681, 5550-5681, 5554-5590, 5605-6137, 5619-6197, 5680-5763, 5680-5773, 5680-5853, 5680-5864, 5680-5869, 5680-5946, 5680-5947, 5680-5974, 5680-6032, 5680-6108, 5680-6197, 5683-5857, 5686-5947, 5686-6190, 5688-6197, 5700-5821, 5723-6197, 5760-6197, 5777-6197, 5792-5913, 5794-6197, 5813-6194, 5837-6145, 5837-6197, 5846-6197, 5868-6197, 5870-6197,

Table 4 (cont.)

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	5890-6073, 5910-6197, 5918-6197, 5923-6197, 5938-6197, 5959-6197, 5965-6197, 5969-6197, 5978-6197, 6023-6131, 6039-6143, 6053-6197, 6092-6197, 6112-6197, 6121-6197, 6133-6197, 6135-6197, 6138-6197
19/ 7156379CB1/ 6367	1-866, 221-866, 354-816, 682-1753, 828-1414, 1000-1569, 1111-1636, 1165-1787, 1194-1787, 1300-1787, 1476-2188, 1525-1936, 1550-1936, 1551-1936, 1612-1936, 1631-1936, 1660-1936, 1672-1936, 1682-1936, 2067-6365, 2171-5224, 2178-2511, 2178-2894, 2317-2833, 2317-2877, 2317-2979, 2317-2995, 2317-3004, 2382-3047, 2561-3391, 2611-3187, 2739-3538, 2741-3538, 2742-3531, 2746-3538, 2764-3538, 3045-3421, 3296-3957, 3345-3957, 3425-3891, 3425-3959, 3425-3966, 4023-4450, 4023-4636, 4856-5187, 4868-5490, 5169-5487, 5252-6250, 5409-5980, 5599-6247, 5620-6223, 5734-6165, 5734-6204, 5734-6233, 5817-6367
20/ 7473626CB1/1615	1-55, 1-597, 1-704, 251-701, 557-766, 612-1276, 785-1279, 826-1247, 933-1377, 1095-1377, 1095-1593, 1114-1280, 1114-1615

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
11	4350981CB1	FTUBTUR01
12	7596315CB1	LUNGFER04
13	71234712CB1	LATRTUT02
14	079370CB1	THYMDIT01
15	2496174CB1	ADRETUE04
16	4097936CB1	ENDANOT01
17	2523646CB1	BRAYDIN03
18	4099073CB1	BRAITUT26
19	7156379CB1	BRAIFER05
20	7473626CB1	BRAIFER06

Table 6

Library	Vector	Library Description
ADRETUE04	PCDNA2.1	This 5 prime biased random primed library was constructed using RNA isolated from adrenal tumor tissue removed from a 52-year-old Caucasian female during a unilateral adrenalectomy. Pathology indicated a pheochromocytoma. Patient history included benign hypertension, depressive disorder, chronic sinusitis, idiopathic proctocolitis, a cataract, and urinary tract infection. Previous surgeries included a vaginal hysterectomy. Patient medications included Procardia (one dose only) and Prozac for 5 years. Family history included secondary Parkinsonism in the father; cerebrovascular disease, secondary Parkinsonism and anxiety state in the mother; and benign hypertension, atherosclerotic coronary artery disease, hyperlipidemia, and brain cancer in the sibling(s).
BRAIFER05	pINCY	Library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation.
BRAIFER06	PCDNA2.1	This random primed library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation. Serologies were negative.
BRAITUT26	pINCY	Library was constructed using RNA isolated from brain tumor tissue removed from the right posterior fossa, occipital convexity of a 70-year-old Caucasian male during cerebral meninges lesion excision. Pathology indicated meningioma. Patient history included a benign colon neoplasm and unspecified personality disorder. Family history included chronic proliferative nephritis, acute myocardial infarction, atherosclerotic coronary artery disease, and chronic proliferative nephritis.
BRAYDIN03	pINCY	This normalized library was constructed from 6.7 million independent clones from a brain tissue library. Starting RNA was made from RNA isolated from diseased hypothalamus tissue removed from a 57-year-old Caucasian male who died from a cerebrovascular accident. Patient history included Huntington's disease and emphysema. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 -hours/round) reannealing hybridization was used. The library was linearized and recircularized to select for insert containing clones.

Table 6 (cont.)

Library	Vector	Library Description
ENDANOT01	PBLUESCRIPT	Library was constructed using RNA isolated from aortic endothelial cell tissue from an explanted heart removed from a male during a heart transplant.
FTUBTUR01	PCDNA2.1	This random primed library was constructed using RNA isolated from fallopian tube tumor tissue removed from an 85-year-old Caucasian female during bilateral salpingo-oophorectomy and hysterectomy. Pathology indicated poorly differentiated mixed endometrioid (80%) and serous (20%) adenocarcinoma, which was confined to the mucosa without mural involvement. Endometrioid carcinoma in situ was also present. Pathology for the associated uterus tumor indicated focal endometrioid adenocarcinoma in situ and moderately differentiated invasive adenocarcinoma arising in an endometrial polyp. Metastatic endometrioid and serous adenocarcinoma was present at the cul-de-sac tumor. Patient history included medullary carcinoma of the thyroid and myocardial infarction.
LATRTUT02	pINCY	Library was constructed using RNA isolated from a myxoma removed from the left atrium of a 43-year-old Caucasian male during annuloplasty. Pathology indicated atrial myxoma. Patient history included pulmonary insufficiency, acute myocardial infarction, atherosclerotic coronary artery disease, hyperlipidemia, and tobacco use. Family history included benign hypertension, acute myocardial infarction, atherosclerotic coronary artery disease, and type II diabetes.
LUNGFER04	PCDNA2.1	This random primed library was constructed using RNA isolated from lung tissue removed from a Caucasian male fetus who died from fetal demise.
THYMDIT01	pINCY	The library was constructed using RNA isolated from diseased thymus tissue removed from a 16-year-old Caucasian female during a total excision of thymus and regional lymph node excision. Pathology indicated thymic follicular hyperplasia. The right lateral thymus showed reactive lymph nodes. A single reactive lymph node was also identified at the inferior thymus margin. The patient presented with myasthenia gravis, malaise, fatigue, dysphagia, severe muscle weakness, and prominent eyes. Patient history included frozen face muscles. Family history included depressive disorder, hepatitis B, myocardial infarction, atherosclerotic coronary artery disease, leukemia, multiple sclerosis, and lupus.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	*ESTs:* Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	*PFAM hits:* Probability value= 1.0E-3 or less *Signal peptide hits:* Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10,

b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10,

c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and

d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method of producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant

polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20,

c) a polynucleotide complementary to a polynucleotide of a),

d) a polynucleotide complementary to a polynucleotide of b), and

e) an RNA equivalent of a)-d).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

19. A method for treating a disease or condition associated with decreased expression of functional CADHP, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional CADHP, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional CADHP, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

a) treating a biological sample containing nucleic acids with the test compound,

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

c) quantifying the amount of hybridization complex, and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A diagnostic test for a condition or disease associated with the expression of CADHP in a biological sample, the method comprising:

a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and

b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

a) a chimeric antibody,

b) a single chain antibody,

c) a Fab fragment,

d) a $F(ab')_2$ fragment, or

e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of CADHP in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of CADHP in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibodies from said animal, and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibody producing cells from the animal,

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,

d) culturing the hybridoma cells, and

e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10 in a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10 from a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

a) labeling the polynucleotides of the sample,

b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides

of the sample under conditions suitable for the formation of a hybridization complex, and

c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

66. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:11.

67. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:12.

68. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:13.

69. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:14.

70. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:15.

71. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:16.

72. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:17.

73. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:18.

74. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:19.

75. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:20.

<110> INCYTE GENOMICS, INC.
DUGGAN, Brendan M.
XU, Yuming
LEE, Ernestine A.
LEE, Sally
LU, Dyung Aina M.
WARREN, Bridget A.
YUE, Henry
GIETZEN, Kimberly J.
HONCHELL, Cynthia D.
BURFORD, Neil
BAUGHN, Mariah R.
TANG, Y. Tom
HILLMAN, Jennifer L.
GANDHI, Ameena R.
KALLICK, Deborah A.
BANDMAN, Olga
GRAUL, Richard C.
WALIA, Narinder K.
LU, Yan
RAMKUMAR, Jayalaxmi
YAO, Monique G.
LAL, Preeti G.

<120> CELL ADHESION PROTEINS

<130> PF-0867 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/256,542; 60/259,604; 60/260,101
<151> 2000-12-18; 2000-12-22; 2001-01-05

<160> 20
<170> PERL Program

<210> 1
<211> 523
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4350981CD1

<400> 1
Met Asp Gly Glu Ala Val Arg Phe Cys Thr Asp Asn Gln Cys Val
1 5 10 15
Ser Leu His Pro Gln Glu Val Asp Ser Val Ala Met Ala Pro Ala
20 25 30
Ala Pro Lys Ile Pro Arg Leu Val Gln Ala Thr Pro Ala Phe Met
35 40 45
Ala Val Thr Leu Val Phe Ser Leu Val Thr Leu Phe Val Val Asp
50 55 60
His His His Phe Gly Arg Glu Ala Glu Met Arg Glu Leu Ile Gln

```
                    65                  70                  75
Thr Phe Lys Gly His Met Glu Asn Ser Ser Ala Trp Val Val Glu
                    80                  85                  90
Ile Gln Met Leu Lys Cys Arg Val Asp Asn Val Asn Ser Gln Leu
                    95                 100                 105
Gln Val Leu Gly Asp His Leu Gly Asn Thr Asn Ala Asp Ile Gln
                   110                 115                 120
Met Val Lys Gly Val Leu Lys Asp Ala Thr Thr Leu Ser Leu Gln
                   125                 130                 135
Thr Gln Met Leu Arg Ser Ser Leu Glu Gly Thr Asn Ala Glu Ile
                   140                 145                 150
Gln Arg Leu Lys Glu Asp Leu Glu Lys Ala Asp Ala Leu Thr Phe
                   155                 160                 165
Gln Thr Leu Asn Phe Leu Lys Ser Ser Leu Glu Asn Thr Ser Ile
                   170                 175                 180
Glu Leu His Val Leu Ser Arg Gly Leu Glu Asn Ala Asn Ser Glu
                   185                 190                 195
Ile Gln Met Leu Asn Ala Ser Leu Glu Thr Ala Asn Thr Gln Ala
                   200                 205                 210
Gln Leu Ala Asn Ser Ser Leu Lys Asn Ala Asn Ala Glu Ile Tyr
                   215                 220                 225
Val Leu Arg Gly His Leu Asp Ser Val Asn Asp Leu Arg Thr Gln
                   230                 235                 240
Asn Gln Val Leu Arg Asn Ser Leu Glu Gly Ala Asn Ala Glu Ile
                   245                 250                 255
Gln Gly Leu Lys Glu Asn Leu Gln Asn Thr Asn Ala Leu Asn Ser
                   260                 265                 270
Gln Thr Gln Ala Phe Ile Lys Ser Ser Phe Asp Asn Thr Ser Ala
                   275                 280                 285
Glu Ile Gln Phe Leu Arg Gly His Leu Glu Arg Ala Gly Asp Glu
                   290                 295                 300
Ile His Val Leu Lys Arg Asp Leu Lys Met Val Thr Ala Gln Thr
                   305                 310                 315
Gln Lys Ala Asn Gly Arg Leu Asp Gln Thr Asp Thr Gln Ile Gln
                   320                 325                 330
Val Phe Lys Ser Glu Met Glu Asn Val Asn Thr Leu Asn Ala Gln
                   335                 340                 345
Ile Gln Val Leu Asn Gly His Met Lys Asn Ala Ser Arg Glu Ile
                   350                 355                 360
Gln Thr Leu Lys Gln Gly Met Lys Asn Ala Ser Ala Leu Thr Ser
                   365                 370                 375
Gln Thr Gln Met Leu Asp Ser Asn Leu Gln Lys Ala Ser Ala Glu
                   380                 385                 390
Ile Gln Arg Leu Arg Gly Asp Leu Glu Asn Thr Lys Ala Leu Thr
                   395                 400                 405
Met Glu Ile Gln Gln Glu Gln Ser Arg Leu Lys Thr Leu His Val
                   410                 415                 420
Val Ile Thr Ser Gln Glu Gln Leu Gln Arg Thr Gln Ser Gln Leu
                   425                 430                 435
Leu Gln Met Val Leu Gln Gly Trp Lys Phe Asn Gly Gly Ser Leu
                   440                 445                 450
Tyr Tyr Phe Ser Ser Val Lys Lys Ser Trp His Glu Ala Glu Gln
                   455                 460                 465
Phe Cys Val Ser Gln Gly Ala His Leu Ala Ser Val Ala Ser Lys
                   470                 475                 480
Glu Glu Gln Ala Phe Leu Val Glu Phe Thr Ser Lys Val Tyr Tyr
```

```
                485                 490                 495
Trp Ile Gly Leu Thr Asp Arg Gly Thr Glu Gly Ser Trp Arg Trp
                500                 505                 510
Thr Asp Gly Thr Pro Phe Asn Ala Ala Gln Asn Lys Ala
                515                 520
```

<210> 2
<211> 1017
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7596315CD1

<400> 2

```
Met Arg Val Phe Leu Leu Cys Ala Tyr Ile Leu Leu Leu Met Val
  1               5                  10                  15
Ser Gln Leu Arg Ala Val Ser Phe Pro Glu Asp Asp Glu Pro Leu
             20                  25                  30
Asn Thr Val Asp Tyr His Tyr Ser Arg Gln Tyr Pro Val Phe Arg
             35                  40                  45
Gly Arg Pro Ser Gly Asn Glu Ser Gln His Arg Leu Asp Phe Gln
             50                  55                  60
Leu Met Leu Lys Ile Arg Asp Thr Leu Tyr Ile Ala Gly Arg Asp
             65                  70                  75
Gln Val Tyr Thr Val Asn Leu Asn Glu Met Pro Lys Thr Glu Val
             80                  85                  90
Ile Pro Asn Lys Lys Leu Thr Trp Arg Ser Arg Gln Gln Asp Arg
             95                 100                 105
Glu Asn Cys Ala Met Lys Gly Lys His Lys Asp Glu Cys His Asn
            110                 115                 120
Phe Ile Lys Val Phe Val Pro Arg Asn Asp Glu Met Val Phe Val
            125                 130                 135
Cys Gly Thr Asn Ala Phe Asn Pro Met Cys Arg Tyr Tyr Arg Leu
            140                 145                 150
Ser Thr Leu Glu Tyr Asp Gly Glu Glu Ile Ser Gly Leu Ala Arg
            155                 160                 165
Cys Pro Phe Asp Ala Arg Gln Thr Asn Val Ala Leu Phe Ala Asp
            170                 175                 180
Gly Lys Leu Tyr Ser Ala Thr Val Ala Asp Phe Leu Ala Ser Asp
            185                 190                 195
Ala Val Ile Tyr Arg Ser Met Gly Asp Gly Ser Ala Leu Arg Thr
            200                 205                 210
Ile Lys Tyr Asp Ser Lys Trp Ile Lys Glu Pro His Phe Leu His
            215                 220                 225
Ala Ile Glu Tyr Gly Asn Tyr Val Tyr Phe Phe Phe Arg Glu Ile
            230                 235                 240
Ala Val Glu His Asn Asn Leu Gly Lys Ala Val Tyr Ser Arg Val
            245                 250                 255
Ala Arg Ile Cys Lys Asn Asp Met Gly Gly Ser Gln Arg Val Leu
            260                 265                 270
Glu Lys His Trp Thr Ser Phe Leu Lys Ala Arg Leu Asn Cys Ser
            275                 280                 285
Val Pro Gly Asp Ser Phe Phe Tyr Phe Asp Val Leu Gln Ser Ile
            290                 295                 300
```

Thr Asp Ile Ile Gln Ile Asn Gly Ile Pro Thr Val Val Gly Val
305 310 315
Phe Thr Thr Gln Leu Asn Ser Ile Pro Gly Ser Ala Val Cys Ala
320 325 330
Phe Ser Met Asp Asp Ile Glu Lys Val Phe Lys Gly Arg Phe Lys
335 340 345
Glu Gln Lys Thr Pro Asp Ser Val Trp Thr Ala Val Pro Glu Asp
350 355 360
Lys Val Pro Lys Pro Arg Pro Gly Cys Cys Ala Lys His Gly Leu
365 370 375
Ala Glu Ala Tyr Lys Thr Ser Ile Asp Phe Pro Asp Glu Thr Leu
380 385 390
Ser Phe Ile Lys Ser His Pro Leu Met Asp Ser Ala Val Pro Pro
395 400 405
Ile Ala Asp Glu Pro Trp Phe Thr Lys Thr Arg Val Arg Tyr Arg
410 415 420
Leu Thr Ala Ile Ser Val Asp His Ser Ala Gly Pro Tyr Gln Asn
425 430 435
Tyr Thr Val Ile Phe Val Gly Ser Glu Ala Gly Met Val Leu Lys
440 445 450
Val Leu Ala Lys Thr Ser Pro Phe Ser Leu Asn Asp Ser Val Leu
455 460 465
Leu Glu Glu Ile Glu Ala Tyr Asn His Ala Lys Cys Asn Ala Glu
470 475 480
Asn Glu Glu Asp Lys Lys Val Ile Ser Leu Gln Leu Asp Lys Asp
485 490 495
His His Ala Leu Tyr Val Ala Phe Ser Ser Cys Ile Ile Arg Ile
500 505 510
Pro Leu Ser Arg Cys Glu Arg Tyr Gly Ser Cys Lys Lys Ser Cys
515 520 525
Ile Ala Ser Arg Asp Pro Tyr Cys Gly Trp Leu Ser Gln Gly Ser
530 535 540
Cys Gly Arg Val Thr Pro Gly Met Leu Ala Glu Gly Tyr Glu Gln
545 550 555
Asp Thr Glu Phe Gly Asn Thr Ala His Leu Gly Asp Cys His Glu
560 565 570
Ile Leu Pro Thr Ser Thr Thr Pro Asp Tyr Lys Ile Phe Gly Gly
575 580 585
Pro Thr Ser Gly Val Arg Trp Glu Val Gln Ser Gly Glu Ser Asn
590 595 600
Gln Met Val His Met Asn Val Leu Ile Thr Cys Val Phe Ala Ala
605 610 615
Phe Val Leu Gly Ala Phe Ile Ala Gly Val Ala Val Tyr Cys Tyr
620 625 630
Arg Asp Met Phe Val Arg Lys Asn Arg Lys Ile His Lys Asp Ala
635 640 645
Glu Ser Ala Gln Ser Cys Thr Asp Ser Ser Gly Ser Phe Ala Lys
650 655 660
Leu Asn Gly Leu Phe Asp Ser Pro Val Lys Glu Tyr Gln Gln Asn
665 670 675
Ile Asp Ser Pro Lys Leu Tyr Ser Asn Leu Leu Thr Ser Arg Lys
680 685 690
Glu Leu Pro Pro Asn Gly Asp Thr Lys Ser Met Val Met Asp His
695 700 705
Arg Gly Gln Pro Pro Glu Leu Ala Ala Leu Pro Thr Pro Glu Ser
710 715 720

```
Thr Pro Val Leu His Gln Lys Thr Leu Gln Ala Met Lys Ser His
                725                 730                 735
Ser Glu Lys Ala His Gly His Gly Ala Ser Arg Lys Glu Thr Pro
                740                 745                 750
Gln Phe Phe Pro Ser Ser Pro Pro Pro His Ser Pro Leu Ser His
                755                 760                 765
Gly His Ile Pro Ser Ala Ile Val Leu Pro Asn Ala Thr His Asp
                770                 775                 780
Tyr Asn Thr Ser Phe Ser Asn Ser Asn Ala His Lys Ala Glu Lys
                785                 790                 795
Lys Leu Gln Asn Ile Asp His Pro Leu Thr Lys Ser Ser Ser Lys
                800                 805                 810
Arg Asp His Arg Arg Ser Val Asp Ser Arg Asn Thr Leu Asn Asp
                815                 820                 825
Leu Leu Lys His Leu Asn Asp Pro Asn Ser Asn Pro Lys Ala Ile
                830                 835                 840
Met Gly Asp Ile Gln Met Ala His Gln Asn Leu Met Leu Asp Pro
                845                 850                 855
Met Gly Ser Met Ser Glu Val Pro Pro Lys Val Pro Asn Arg Glu
                860                 865                 870
Ala Ser Leu Tyr Ser Pro Pro Ser Thr Leu Pro Arg Asn Ser Pro
                875                 880                 885
Thr Lys Arg Val Asp Val Pro Thr Thr Pro Gly Val Pro Met Thr
                890                 895                 900
Ser Leu Glu Arg Gln Arg Gly Tyr His Lys Asn Ser Ser Gln Arg
                905                 910                 915
His Ser Ile Ser Ala Met Pro Lys Asn Leu Asn Ser Pro Asn Gly
                920                 925                 930
Val Leu Leu Ser Arg Gln Pro Ser Met Asn Arg Gly Gly Tyr Met
                935                 940                 945
Pro Thr Pro Thr Gly Ala Lys Val Asp Tyr Ile Gln Gly Thr Pro
                950                 955                 960
Val Ser Val His Leu Gln Pro Ser Leu Ser Arg Gln Ser Ser Tyr
                965                 970                 975
Thr Ser Asn Gly Thr Leu Pro Arg Thr Gly Leu Lys Arg Thr Pro
                980                 985                 990
Ser Leu Lys Pro Asp Val Pro Pro Lys Pro Ser Phe Val Pro Gln
                995                1000                1005
Thr Pro Ser Val Arg Pro Leu Asn Lys Tyr Thr Tyr
               1010                1015

<210> 3
<211> 561
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 71234712CD1

<400> 3
Met Ala Ser Arg Ala Val Val Arg Ala Arg Arg Cys Pro Gln Cys
  1               5                  10                  15
Pro Gln Val Arg Ala Ala Ala Ala Ala Pro Ala Trp Ala Ala Leu
                20                  25                  30
Pro Leu Ser Arg Ser Leu Pro Pro Cys Ser Asn Ser Ser Ser Phe
```

```
                35                  40                  45
Ser Met Pro Leu Phe Leu Leu Leu Leu Leu Val Leu Leu Leu Leu
                50                  55                  60
Leu Glu Asp Ala Gly Ala Gln Gln Gly Asp Gly Cys Gly His Thr
                65                  70                  75
Val Leu Gly Pro Glu Ser Gly Thr Leu Thr Ser Ile Asn Tyr Pro
                80                  85                  90
Gln Thr Tyr Pro Asn Ser Thr Val Cys Glu Trp Glu Ile Arg Val
                95                 100                 105
Lys Met Gly Glu Arg Val Arg Ile Lys Phe Gly Asp Phe Asp Ile
               110                 115                 120
Glu Asp Ser Asp Ser Cys His Phe Asn Tyr Leu Arg Ile Tyr Asn
               125                 130                 135
Gly Ile Gly Val Ser Arg Thr Glu Ile Gly Lys Tyr Cys Gly Leu
               140                 145                 150
Gly Leu Gln Met Asn His Ser Ile Glu Ser Lys Gly Asn Glu Ile
               155                 160                 165
Thr Leu Leu Phe Met Ser Gly Ile His Val Ser Gly Arg Gly Phe
               170                 175                 180
Leu Ala Ser Tyr Ser Val Ile Asp Lys Gln Asp Leu Ile Thr Cys
               185                 190                 195
Leu Asp Thr Ala Ser Asn Phe Leu Glu Pro Glu Phe Ser Lys Tyr
               200                 205                 210
Cys Pro Ala Gly Cys Leu Leu Pro Phe Ala Glu Ile Ser Gly Thr
               215                 220                 225
Ile Pro His Gly Tyr Arg Asp Ser Ser Pro Leu Cys Met Ala Gly
               230                 235                 240
Val His Ala Gly Val Val Ser Asn Thr Leu Gly Gly Gln Ile Ser
               245                 250                 255
Val Val Ile Ser Lys Gly Ile Pro Tyr Tyr Glu Ser Ser Leu Ala
               260                 265                 270
Asn Asn Val Thr Ser Val Val Gly His Leu Ser Thr Ser Leu Phe
               275                 280                 285
Thr Phe Lys Thr Ser Gly Cys Tyr Gly Thr Leu Gly Met Glu Ser
               290                 295                 300
Gly Val Ile Ala Asp Pro Gln Ile Thr Ala Ser Ser Val Leu Glu
               305                 310                 315
Trp Thr Asp His Thr Gly Gln Glu Asn Ser Trp Lys Pro Lys Lys
               320                 325                 330
Ala Arg Leu Lys Lys Pro Gly Pro Pro Trp Ala Ala Phe Ala Thr
               335                 340                 345
Asp Glu Tyr Gln Trp Leu Gln Ile Asp Leu Asn Lys Glu Lys Lys
               350                 355                 360
Ile Thr Gly Ile Ile Thr Thr Gly Ser Thr Met Val Glu His Asn
               365                 370                 375
Tyr Tyr Val Ser Ala Tyr Arg Ile Leu Tyr Ser Asp Asp Gly Gln
               380                 385                 390
Lys Trp Thr Val Tyr Arg Glu Pro Gly Val Glu Gln Asp Lys Ile
               395                 400                 405
Phe Gln Gly Asn Lys Asp Tyr His Gln Asp Val Arg Asn Asn Phe
               410                 415                 420
Leu Pro Pro Ile Ile Ala Arg Phe Ile Arg Val Asn Pro Thr Gln
               425                 430                 435
Trp Gln Gln Lys Ile Ala Met Lys Met Glu Leu Leu Gly Cys Gln
               440                 445                 450
Phe Ile Pro Lys Gly Arg Pro Pro Lys Leu Thr Gln Pro Pro Pro
```

```
                455                 460                 465
Pro Arg Asn Ser Asn Asp Leu Lys Asn Thr Thr Ala Pro Pro Lys
                470                 475                 480
Ile Ala Lys Gly Arg Ala Pro Lys Phe Thr Gln Pro Leu Gln Pro
                485                 490                 495
Arg Ser Ser Asn Glu Phe Pro Ala Gln Thr Glu Gln Thr Thr Ala
                500                 505                 510
Ser Pro Asp Ile Arg Asn Thr Thr Val Thr Pro Asn Val Thr Lys
                515                 520                 525
Asp Val Ala Leu Ala Ala Val Leu Val Pro Val Leu Val Met Val
                530                 535                 540
Leu Thr Thr Leu Ile Leu Ile Leu Val Cys Ala Trp His Trp Arg
                545                 550                 555
Asn Arg Leu Val His Asn
                560
```

<210> 4
<211> 439
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 079370CD1

<400> 4

```
Met Val Pro Ser Ser Pro Arg Ala Leu Phe Leu Leu Leu Leu Ile
  1               5                  10                  15
Leu Ala Cys Pro Glu Pro Arg Ala Ser Gln Asn Cys Leu Ser Lys
                20                  25                  30
Gln Gln Leu Leu Ser Ala Ile Arg Gln Leu Gln Gln Leu Leu Lys
                35                  40                  45
Gly Gln Glu Thr Arg Phe Ala Glu Gly Ile Arg His Met Lys Ser
                50                  55                  60
Arg Leu Ala Ala Leu Gln Asn Ser Val Gly Arg Val Gly Pro Asp
                65                  70                  75
Ala Leu Pro Val Ser Cys Pro Ala Leu Asn Thr Pro Ala Asp Gly
                80                  85                  90
Arg Lys Phe Gly Ser Lys Tyr Leu Val Asp His Glu Val His Phe
                95                 100                 105
Thr Cys Asn Pro Gly Phe Arg Leu Val Gly Pro Ser Ser Val Val
               110                 115                 120
Cys Leu Pro Asn Gly Thr Trp Thr Gly Glu Gln Pro His Cys Arg
               125                 130                 135
Gly Ile Ser Glu Cys Ser Ser Gln Pro Cys Gln Asn Gly Gly Thr
               140                 145                 150
Cys Val Glu Gly Val Asn Gln Tyr Arg Cys Ile Cys Pro Pro Gly
               155                 160                 165
Arg Thr Gly Asn Arg Cys Gln His Gln Ala Gln Thr Ala Ala Pro
               170                 175                 180
Glu Gly Ser Val Ala Gly Asp Ser Ala Phe Ser Arg Ala Pro Arg
               185                 190                 195
Cys Ala Gln Val Glu Arg Ala Gln His Cys Ser Cys Glu Ala Gly
               200                 205                 210
Phe His Leu Ser Gly Ala Ala Gly Asp Ser Val Cys Gln Asp Val
               215                 220                 225
```

```
Asn Glu Cys Glu Leu Tyr Gly Gln Glu Gly Arg Pro Arg Leu Cys
                230                 235                 240
Met His Ala Cys Val Asn Thr Pro Gly Ser Tyr Arg Cys Thr Cys
                245                 250                 255
Pro Gly Gly Tyr Arg Thr Leu Ala Asp Gly Lys Ser Cys Glu Asp
                260                 265                 270
Val Asp Glu Cys Val Gly Leu Gln Pro Val Cys Pro Gln Gly Thr
                275                 280                 285
Thr Cys Ile Asn Thr Gly Gly Ser Phe Gln Cys Val Ser Pro Glu
                290                 295                 300
Cys Pro Glu Gly Ser Gly Asn Val Ser Tyr Val Lys Thr Ser Pro
                305                 310                 315
Phe Gln Cys Glu Arg Asn Pro Cys Pro Met Asp Ser Arg Pro Cys
                320                 325                 330
Arg His Leu Pro Lys Thr Ile Ser Phe His Tyr Leu Ser Leu Pro
                335                 340                 345
Ser Asn Leu Lys Thr Pro Ile Thr Leu Phe Arg Met Ala Thr Ala
                350                 355                 360
Ser Ala Pro Gly Arg Ala Gly Pro Asn Ser Leu Arg Phe Gly Ile
                365                 370                 375
Val Gly Gly Asn Ser Arg Gly His Phe Val Met Gln Arg Ser Asp
                380                 385                 390
Arg Gln Thr Gly Asp Leu Ile Leu Val Gln Asn Leu Glu Gly Pro
                395                 400                 405
Gln Thr Leu Glu Val Asp Val Asp Met Ser Glu Tyr Leu Asp Arg
                410                 415                 420
Ser Phe Gln Ala Asn His Val Ser Lys Val Thr Ile Phe Val Ser
                425                 430                 435
Pro Tyr Asp Phe
```

<210> 5
<211> 160
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2496174CD1

<400> 5

```
Met Val Leu Ala Gly Pro Leu Ala Val Ser Leu Leu Leu Pro Ser
  1               5                  10                  15
Leu Thr Leu Leu Val Ser His Leu Ser Ser Ser Gln Asp Val Ser
                20                  25                  30
Ser Glu Pro Ser Ser Glu Gln Gln Leu Cys Ala Leu Ser Lys His
                35                  40                  45
Pro Thr Val Ala Phe Glu Asp Leu Gln Pro Trp Val Ser Asn Phe
                50                  55                  60
Thr Tyr Pro Gly Ala Arg Asp Phe Ser Gln Leu Ala Leu Asp Pro
                65                  70                  75
Ser Gly Asn Gln Leu Ile Val Gly Ala Arg Asn Tyr Leu Phe Arg
                80                  85                  90
Leu Ser Leu Ala Asn Val Ser Leu Leu Gln Glu Asp Thr Gly Asp
                95                 100                 105
Val Phe His Gln Asn Lys Arg Ile Asn Gln Glu Arg Gly Lys His
```

```
                110                 115                 120
Ala Ile Arg Lys Ala Gly Glu Glu Arg Arg Pro Gln Ser Gly Pro
                125                 130                 135
Pro Val Arg Thr Arg Ala Ala Pro Ala Lys Ala Lys Gly Arg Leu
                140                 145                 150
Arg Arg Ser Val Arg Thr Thr Cys Glu Ser
                155                 160

<210> 6
<211> 994
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4097936CD1

<400> 6
Met Ser Pro Pro Leu Cys Pro Leu Leu Leu Leu Ala Val Gly Leu
  1                 5                  10                  15
Arg Leu Ala Gly Thr Leu Asn Pro Ser Asp Pro Asn Thr Cys Ser
                 20                  25                  30
Phe Trp Glu Ser Phe Thr Thr Thr Thr Lys Glu Ser His Ser Arg
                 35                  40                  45
Pro Phe Ser Leu Leu Pro Ser Glu Pro Cys Glu Arg Pro Trp Glu
                 50                  55                  60
Gly Pro His Thr Cys Pro Gln Pro Thr Val Val Tyr Arg Thr Val
                 65                  70                  75
Tyr Arg Gln Val Val Lys Thr Asp His Arg Gln Arg Leu Gln Cys
                 80                  85                  90
Cys His Gly Phe Tyr Glu Ser Arg Gly Phe Cys Val Pro Leu Cys
                 95                 100                 105
Ala Gln Glu Cys Val His Gly Arg Cys Val Ala Pro Asn Gln Cys
                110                 115                 120
Gln Cys Val Pro Gly Trp Arg Gly Asp Asp Cys Ser Ser Glu Cys
                125                 130                 135
Ala Pro Gly Met Trp Gly Pro Gln Cys Asp Lys Pro Cys Ser Cys
                140                 145                 150
Gly Asn Asn Ser Ser Cys Asp Pro Lys Ser Gly Val Cys Ser Cys
                155                 160                 165
Pro Ser Gly Leu Gln Pro Pro Asn Cys Leu Gln Pro Cys Thr Pro
                170                 175                 180
Gly Tyr Tyr Gly Pro Ala Cys Gln Phe Arg Cys Gln Cys His Gly
                185                 190                 195
Ala Pro Cys Asp Pro Gln Thr Gly Ala Cys Phe Cys Pro Ala Glu
                200                 205                 210
Arg Thr Gly Pro Ser Cys Asp Val Ser Cys Ser Gln Gly Thr Ser
                215                 220                 225
Gly Phe Phe Cys Pro Ser Thr His Pro Cys Gln Asn Gly Gly Val
                230                 235                 240
Phe Gln Thr Pro Gln Gly Ser Cys Ser Cys Pro Pro Gly Trp Met
                245                 250                 255
Gly Thr Ile Cys Ser Leu Pro Cys Pro Glu Gly Phe His Gly Pro
                260                 265                 270
Asn Cys Ser Gln Glu Cys Arg Cys His Asn Gly Gly Leu Cys Asp
                275                 280                 285
```

```
Arg Phe Thr Gly Gln Cys Arg Cys Ala Pro Gly Tyr Thr Gly Asp
                290                 295                 300
Arg Cys Arg Glu Glu Cys Pro Val Gly Arg Phe Gly Gln Asp Cys
                305                 310                 315
Ala Glu Thr Cys Asp Cys Ala Pro Asp Ala Arg Cys Phe Pro Ala
                320                 325                 330
Asn Gly Ala Cys Leu Cys Glu His Gly Phe Thr Gly Asp Arg Cys
                335                 340                 345
Thr Asp Arg Leu Cys Pro Asp Gly Phe Tyr Gly Leu Ser Cys Gln
                350                 355                 360
Ala Pro Cys Thr Cys Asp Arg Glu His Ser Leu Ser Cys His Pro
                365                 370                 375
Met Asn Gly Glu Cys Ser Cys Leu Pro Gly Trp Ala Gly Leu His
                380                 385                 390
Cys Asn Glu Ser Cys Pro Gln Asp Thr His Gly Pro Gly Cys Gln
                395                 400                 405
Glu His Cys Leu Cys Leu His Gly Gly Val Cys Gln Ala Thr Ser
                410                 415                 420
Gly Leu Cys Gln Cys Ala Pro Gly Tyr Thr Gly Pro His Cys Ala
                425                 430                 435
Ser Leu Cys Pro Pro Asp Thr Tyr Gly Val Asn Cys Ser Ala Arg
                440                 445                 450
Cys Ser Cys Glu Asn Ala Ile Ala Cys Ser Pro Ile Asp Gly Glu
                455                 460                 465
Cys Val Cys Lys Glu Gly Trp Gln Arg Gly Asn Cys Ser Val Pro
                470                 475                 480
Cys Pro Pro Gly Thr Trp Gly Phe Ser Cys Asn Ala Ser Cys Gln
                485                 490                 495
Cys Ala His Glu Ala Val Cys Ser Pro Gln Thr Gly Ala Cys Thr
                500                 505                 510
Cys Thr Pro Gly Trp His Gly Ala His Cys Gln Leu Pro Cys Pro
                515                 520                 525
Lys Gly Gln Phe Gly Glu Gly Cys Ala Ser Arg Cys Asp Cys Asp
                530                 535                 540
His Ser Asp Gly Cys Asp Pro Val His Gly Arg Cys Gln Cys Gln
                545                 550                 555
Ala Gly Trp Met Gly Ala Arg Cys His Leu Ser Cys Pro Glu Gly
                560                 565                 570
Leu Trp Gly Val Asn Cys Ser Asn Thr Cys Thr Cys Lys Asn Gly
                575                 580                 585
Gly Thr Cys Leu Pro Glu Asn Gly Asn Cys Val Cys Ala Pro Gly
                590                 595                 600
Phe Arg Gly Pro Ser Cys Gln Arg Ser Cys Gln Pro Gly Arg Tyr
                605                 610                 615
Gly Lys Arg Cys Val Pro Cys Lys Cys Ala Asn His Ser Phe Cys
                620                 625                 630
His Pro Ser Asn Gly Thr Cys Tyr Cys Leu Ala Gly Trp Thr Gly
                635                 640                 645
Pro Asp Cys Ser Gln Arg Cys Pro Leu Gly Thr Phe Gly Ala Asn
                650                 655                 660
Cys Ser Gln Pro Cys Gln Cys Gly Pro Gly Glu Lys Cys His Pro
                665                 670                 675
Glu Thr Gly Ala Cys Val Cys Pro Pro Gly His Ser Gly Ala Pro
                680                 685                 690
Cys Arg Ile Gly Ile Gln Glu Pro Phe Thr Val Met Pro Thr Thr
                695                 700                 705
```

```
Pro Val Ala Tyr Asn Ser Leu Gly Ala Val Ile Gly Ile Ala Val
                710                 715                 720
Leu Gly Ser Leu Val Val Ala Leu Val Ala Leu Phe Ile Gly Tyr
                725                 730                 735
Arg His Trp Gln Lys Gly Lys Glu His His His Leu Ala Val Ala
                740                 745                 750
Tyr Ser Ser Gly Arg Leu Asp Gly Ser Glu Tyr Val Met Pro Asp
                755                 760                 765
Val Pro Pro Ser Tyr Ser His Tyr Tyr Ser Asn Pro Ser Tyr His
                770                 775                 780
Thr Leu Ser Gln Cys Ser Pro Asn Pro Pro Pro Pro Asn Lys Val
                785                 790                 795
Pro Gly Pro Leu Phe Ala Ser Leu Gln Lys Pro Glu Arg Pro Gly
                800                 805                 810
Gly Ala Gln Gly His Asp Asn His Thr Thr Leu Pro Ala Asp Trp
                815                 820                 825
Lys His Arg Arg Glu Pro Pro Pro Gly Pro Leu Asp Arg Gly Ser
                830                 835                 840
Ser Arg Leu Asp Arg Ser Tyr Ser Tyr Ser Tyr Ser Asn Gly Pro
                845                 850                 855
Gly Pro Phe Tyr Asn Lys Gly Leu Ile Ser Glu Glu Glu Leu Gly
                860                 865                 870
Ala Ser Val Ala Ser Leu Ser Ser Glu Asn Pro Tyr Ala Thr Ile
                875                 880                 885
Arg Asp Leu Pro Ser Leu Pro Gly Gly Pro Arg Glu Ser Ser Tyr
                890                 895                 900
Met Glu Met Lys Gly Pro Pro Ser Gly Ser Pro Pro Arg Gln Pro
                905                 910                 915
Pro Gln Phe Trp Asp Ser Gln Arg Arg Arg Gln Pro Gln Pro Gln
                920                 925                 930
Arg Asp Ser Gly Thr Tyr Glu Gln Pro Ser Pro Leu Ile His Asp
                935                 940                 945
Arg Asp Ser Val Gly Ser Gln Pro Pro Leu Pro Pro Gly Leu Pro
                950                 955                 960
Pro Gly His Tyr Asp Ser Pro Lys Asn Ser His Ile Pro Gly His
                965                 970                 975
Tyr Asp Leu Pro Pro Val Arg His Pro Pro Ser Pro Pro Leu Arg
                980                 985                 990
Arg Gln Asp Arg
```

<210> 7
<211> 987
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2523646CD1

<400> 7

```
Met Lys Met Thr Arg Pro Arg Val Trp Leu Ala Glu Gly Cys Arg
  1               5                  10                  15
Glu Trp Ala Leu Arg Asp Ser Ala Leu Met Ala Gln Leu Leu Arg
                20                  25                  30
Thr Gly Ser Pro Leu Tyr Leu Leu Cys Ser His Pro Gln Asn Thr
```

```
                35                  40                  45
Pro Val Gly Thr Pro Ile Phe Ile Val Asn Ala Thr Asp Pro Asp
            50                  55                  60
Leu Gly Ala Gly Gly Ser Val Leu Tyr Ser Phe Gln Pro Pro Ser
        65                  70                  75
Gln Phe Phe Ala Ile Asp Ser Ala Arg Gly Ile Val Thr Val Ile
    80                  85                  90
Arg Glu Leu Asp Tyr Glu Thr Thr Gln Ala Tyr Gln Leu Thr Val
                95                  100                 105
Asn Ala Thr Asp Gln Asp Lys Thr Arg Pro Leu Ser Thr Leu Ala
            110                 115                 120
Asn Leu Ala Ile Ile Ile Thr Asp Val Gln Asp Met Asp Pro Ile
        125                 130                 135
Phe Ile Asn Leu Pro Tyr Ser Thr Asn Ile Tyr Glu His Ser Pro
    140                 145                 150
Pro Gly Thr Thr Val Arg Ile Ile Thr Ala Ile Asp Gln Asp Lys
155                 160                 165
Gly Arg Pro Arg Gly Ile Gly Tyr Thr Ile Val Ser Gly Asn Thr
            170                 175                 180
Asn Ser Ile Phe Ala Leu Asp Tyr Ile Ser Gly Val Leu Thr Leu
                185                 190                 195
Asn Gly Leu Leu Asp Arg Glu Asn Pro Leu Tyr Ser His Gly Phe
            200                 205                 210
Ile Leu Thr Val Lys Gly Thr Glu Leu Asn Asp Asp Arg Thr Pro
        215                 220                 225
Ser Asp Ala Thr Val Thr Thr Thr Phe Asn Ile Leu Val Ile Asp
    230                 235                 240
Ile Asn Asp Asn Ala Pro Glu Phe Asn Ser Ser Glu Tyr Ser Val
245                 250                 255
Ala Ile Thr Glu Leu Ala Gln Val Gly Phe Ala Leu Pro Leu Phe
            260                 265                 270
Ile Gln Val Val Asp Lys Asp Glu Asn Leu Gly Leu Asn Ser Met
        275                 280                 285
Phe Glu Val Tyr Leu Val Gly Asn Asn Ser His His Phe Ile Ile
    290                 295                 300
Ser Pro Thr Ser Val Gln Gly Lys Ala Asp Ile Arg Ile Arg Val
305                 310                 315
Ala Ile Pro Leu Asp Tyr Glu Thr Val Asp Arg Tyr Asp Phe Asp
            320                 325                 330
Leu Phe Ala Asn Glu Ser Val Pro Asp His Val Gly Tyr Ala Lys
        335                 340                 345
Val Lys Ile Thr Leu Ile Asn Glu Asn Asp Asn Arg Pro Ile Phe
    350                 355                 360
Ser Gln Pro Leu Tyr Asn Ile Ser Leu Tyr Glu Asn Val Thr Val
365                 370                 375
Gly Thr Ser Val Leu Thr Val Leu Ala Thr Asp Asn Asp Ala Gly
            380                 385                 390
Thr Phe Gly Glu Val Ser Tyr Phe Phe Ser Asp Asp Pro Asp Arg
        395                 400                 405
Phe Ser Leu Asp Lys Asp Thr Gly Leu Ile Met Leu Ile Ala Arg
    410                 415                 420
Leu Asp Tyr Glu Leu Ile Gln Arg Phe Thr Leu Thr Ile Ile Ala
425                 430                 435
Arg Asp Gly Gly Gly Glu Glu Thr Thr Gly Arg Val Arg Ile Asn
            440                 445                 450
Val Leu Asp Val Asn Asp Asn Val Pro Thr Phe Gln Lys Asp Ala
```

```
                455                 460                 465
Tyr Val Gly Ala Leu Arg Glu Asn Glu Pro Ser Val Thr Gln Leu
                470                 475                 480
Val Arg Leu Arg Ala Thr Asp Glu Asp Ser Pro Pro Asn Asn Gln
                485                 490                 495
Ile Thr Tyr Ser Ile Val Ser Ala Ser Ala Phe Gly Ser Tyr Phe
                500                 505                 510
Asp Ile Ser Leu Tyr Glu Gly Tyr Gly Val Ile Ser Val Ser Arg
                515                 520                 525
Pro Leu Asp Tyr Glu Gln Ile Ser Asn Gly Leu Ile Tyr Leu Thr
                530                 535                 540
Val Met Ala Met Asp Ala Gly Asn Pro Pro Leu Asn Ser Thr Val
                545                 550                 555
Pro Val Thr Ile Glu Val Phe Asp Glu Asn Asp Asn Pro Pro Thr
                560                 565                 570
Phe Ser Lys Pro Ala Tyr Phe Val Ser Val Val Glu Asn Ile Met
                575                 580                 585
Ala Gly Ala Thr Val Leu Phe Leu Asn Ala Thr Asp Leu Asp Arg
                590                 595                 600
Ser Arg Glu Tyr Gly Gln Glu Ser Ile Ile Tyr Ser Leu Glu Gly
                605                 610                 615
Ser Thr Gln Phe Arg Ile Asn Ala Arg Ser Gly Glu Ile Thr Thr
                620                 625                 630
Thr Ser Leu Leu Asp Arg Glu Thr Lys Ser Glu Tyr Ile Leu Ile
                635                 640                 645
Val Arg Ala Val Asp Gly Gly Val Gly His Asn Gln Lys Thr Gly
                650                 655                 660
Ile Ala Thr Val Asn Ile Thr Leu Leu Asp Ile Asn Asp Asn His
                665                 670                 675
Pro Thr Trp Lys Asp Ala Pro Tyr Tyr Ile Asn Leu Val Glu Met
                680                 685                 690
Thr Pro Pro Asp Ser Asp Val Thr Thr Val Val Ala Val Asp Pro
                695                 700                 705
Asp Leu Gly Glu Asn Gly Thr Leu Val Tyr Ser Ile Gln Pro Pro
                710                 715                 720
Asn Lys Phe Tyr Ser Leu Asn Ser Thr Thr Gly Lys Ile Arg Thr
                725                 730                 735
Thr His Ala Met Leu Asp Arg Glu Asn Pro Asp Pro His Glu Ala
                740                 745                 750
Glu Leu Met Arg Lys Ile Val Val Ser Val Thr Asp Cys Gly Arg
                755                 760                 765
Pro Pro Leu Lys Ala Thr Ser Ser Ala Thr Val Phe Val Asn Leu
                770                 775                 780
Leu Asp Leu Asn Asp Asn Asp Pro Thr Phe Gln Asn Leu Pro Phe
                785                 790                 795
Val Ala Glu Val Leu Glu Gly Ile Pro Ala Gly Val Ser Ile Tyr
                800                 805                 810
Gln Val Val Ala Ile Asp Leu Asp Glu Gly Leu Asn Gly Leu Val
                815                 820                 825
Ser Tyr Arg Met Pro Val Gly Met Pro Arg Met Asp Phe Leu Ile
                830                 835                 840
Asn Ser Ser Ser Gly Val Val Val Thr Thr Thr Glu Leu Asp Arg
                845                 850                 855
Glu Arg Ile Ala Glu Tyr Gln Leu Arg Val Val Ala Ser Asp Ala
                860                 865                 870
Gly Thr Pro Thr Lys Ser Ser Thr Ser Thr Leu Thr Ile His Gly
```

```
                875                 880                 885
Cys Ser Glu Gly Cys Met Trp Ser Cys Met Gly Ser Thr Gln His
                890                 895                 900
Gly Leu Gly Thr Leu Asp Lys Leu Val Asn Val Leu Asp Val Asn
                905                 910                 915
Asp Glu Thr Pro Thr Phe Phe Pro Ala Val Tyr Asn Val Ser Val
                920                 925                 930
Ser Glu Asp Val Pro Arg Glu Phe Arg Val Val Trp Leu Asn Cys
                935                 940                 945
Thr Asp Asn Asp Val Gly Leu Asn Ala Glu Leu Ser Tyr Phe Ile
                950                 955                 960
Thr Gly Ala Ala Pro Ala Ser Ala His Leu Cys Arg Pro Pro Gly
                965                 970                 975
Ala Leu Pro Pro Pro Leu Pro Asp Gly Gln Pro Asp
                980                 985
```

<210> 8
<211> 1619
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4099073CD1

<400> 8

```
Met Phe Gln Gln Pro His Tyr Glu Val Leu Leu Asp Glu Gly Pro
  1               5                  10                  15
Asp Thr Leu Asn Thr Ser Leu Ile Thr Ile Gln Ala Leu Asp Leu
                20                  25                  30
Asp Glu Gly Pro Asn Gly Thr Val Thr Tyr Ala Ile Val Ala Gly
                35                  40                  45
Asn Ile Val Asn Thr Phe Arg Ile Asp Arg His Met Gly Val Ile
                50                  55                  60
Thr Ala Ala Lys Glu Leu Asp Tyr Glu Ile Ser His Gly Arg Tyr
                65                  70                  75
Thr Leu Ile Val Thr Ala Thr Asp Gln Cys Pro Ile Leu Ser His
                80                  85                  90
Arg Leu Thr Ser Thr Thr Thr Val Leu Val Asn Val Asn Asp Ile
                95                 100                 105
Asn Asp Asn Val Pro Thr Phe Pro Arg Asp Tyr Glu Gly Pro Phe
               110                 115                 120
Glu Val Thr Glu Gly Gln Pro Gly Pro Arg Val Trp Thr Phe Leu
               125                 130                 135
Ala His Asp Arg Asp Ser Gly Pro Asn Gly Gln Val Glu Tyr Ser
               140                 145                 150
Ile Met Asp Gly Asp Pro Leu Gly Glu Phe Val Ile Ser Pro Val
               155                 160                 165
Glu Gly Val Leu Arg Val Arg Lys Asp Val Glu Leu Asp Arg Glu
               170                 175                 180
Thr Ile Ala Phe Tyr Asn Leu Thr Ile Cys Ala Arg Asp Arg Gly
               185                 190                 195
Met Pro Pro Leu Ser Ser Thr Met Leu Val Gly Ile Arg Val Leu
               200                 205                 210
Asp Ile Asn Asp Asn Asp Pro Val Leu Leu Asn Leu Pro Met Asn
               215                 220                 225
```

```
Ile Thr Ile Ser Glu Asn Ser Pro Val Ser Ser Phe Val Ala His
                230                 235                 240
Val Leu Ala Ser Asp Ala Asp Ser Gly Cys Asn Ala Arg Leu Thr
                245                 250                 255
Phe Asn Ile Thr Ala Gly Asn Arg Glu Arg Ala Phe Phe Ile Asn
                260                 265                 270
Ala Thr Thr Gly Ile Val Thr Val Asn Arg Pro Leu Asp Arg Glu
                275                 280                 285
Arg Ile Pro Glu Tyr Lys Leu Thr Ile Ser Val Lys Asp Asn Pro
                290                 295                 300
Glu Asn Pro Arg Ile Ala Arg Arg Asp Tyr Asp Leu Leu Leu Ile
                305                 310                 315
Phe Leu Ser Asp Glu Asn Asp Asn His Pro Leu Phe Thr Lys Ser
                320                 325                 330
Thr Tyr Gln Ala Glu Val Met Glu Asn Ser Pro Ala Gly Thr Pro
                335                 340                 345
Leu Thr Val Leu Asn Gly Pro Ile Leu Ala Leu Asp Ala Asp Gln
                350                 355                 360
Asp Ile Tyr Ala Val Val Thr Tyr Gln Leu Leu Gly Ala Gln Ser
                365                 370                 375
Gly Leu Phe Asp Ile Asn Ser Ser Thr Gly Phe Ser Val Leu Gln
                380                 385                 390
Val Thr Ala Thr Asp Glu Asp Ser Gly Leu Asn Gly Glu Leu Val
                395                 400                 405
Tyr Arg Ile Glu Ala Gly Ala Gln Asp Arg Phe Leu Ile His Leu
                410                 415                 420
Val Thr Gly Val Ile Arg Val Gly Asn Ala Thr Ile Asp Arg Glu
                425                 430                 435
Glu Gln Glu Ser Tyr Arg Leu Thr Val Val Ala Thr Asp Arg Gly
                440                 445                 450
Thr Val Pro Leu Ser Gly Thr Ala Ile Val Thr Ile Leu Ile Asp
                455                 460                 465
Asp Ile Asn Asp Ser Arg Pro Glu Phe Leu Asn Pro Ile Gln Thr
                470                 475                 480
Val Ser Val Leu Glu Ser Ala Glu Pro Gly Thr Val Ile Ala Asn
                485                 490                 495
Ile Thr Ala Ile Asp His Asp Leu Asn Pro Lys Leu Glu Tyr His
                500                 505                 510
Ile Val Gly Ile Val Ala Lys Asp Asp Thr Asp Arg Leu Val Pro
                515                 520                 525
Asn Gln Glu Asp Ala Phe Ala Val Asn Ile Asn Thr Gly Ser Val
                530                 535                 540
Met Val Lys Ser Pro Met Asn Arg Glu Leu Val Ala Thr Tyr Glu
                545                 550                 555
Val Thr Leu Ser Val Ile Asp Asn Ala Ser Asp Leu Pro Glu Arg
                560                 565                 570
Ser Val Ser Val Pro Asn Ala Lys Leu Thr Val Asn Val Leu Asp
                575                 580                 585
Val Asn Asp Asn Thr Pro Gln Phe Lys Pro Phe Gly Ile Thr Tyr
                590                 595                 600
Tyr Met Glu Arg Ile Leu Glu Gly Ala Thr Pro Gly Thr Thr Leu
                605                 610                 615
Ile Ala Val Ala Ala Val Asp Pro Asp Lys Gly Leu Asn Gly Leu
                620                 625                 630
Val Thr Tyr Thr Leu Leu Asp Leu Val Pro Pro Gly Tyr Val Gln
                635                 640                 645
```

```
Leu Glu Asp Ser Ser Ala Gly Lys Val Ile Ala Asn Gln Thr Val
                650                 655                 660
Asp Tyr Glu Glu Val His Trp Leu Asn Phe Thr Val Arg Ala Ser
                665                 670                 675
Asp Asn Gly Ser Leu Pro Arg Ala Ala Glu Ile Pro Val Tyr Leu
                680                 685                 690
Glu Ile Val Asp Ile Asn Asp Asn Asn Pro Ile Phe Asp Gln Pro
                695                 700                 705
Ser Tyr Gln Glu Ala Val Phe Glu Asp Val Pro Val Gly Thr Ile
                710                 715                 720
Ile Leu Thr Val Thr Ala Thr Asp Ala Asp Ser Gly Asn Phe Ala
                725                 730                 735
Leu Ile Glu Tyr Ser Leu Gly Asp Gly Glu Ser Lys Phe Ala Ile
                740                 745                 750
Asn Pro Thr Thr Gly Asp Ile Tyr Val Leu Ser Ser Leu Asp Arg
                755                 760                 765
Glu Lys Lys Asp His Tyr Ile Leu Thr Ala Leu Ala Lys Asp Asn
                770                 775                 780
Pro Gly Asp Val Ala Ser Asn Arg Arg Glu Asn Ser Val Gln Val
                785                 790                 795
Val Ile Gln Val Leu Asp Val Asn Asp Cys Arg Pro Gln Phe Ser
                800                 805                 810
Lys Pro Gln Phe Ser Thr Ser Val Tyr Glu Asn Glu Pro Ala Gly
                815                 820                 825
Thr Ser Val Ile Thr Met Met Ala Thr Asp Gln Asp Glu Gly Pro
                830                 835                 840
Asn Gly Glu Leu Thr Tyr Ser Leu Glu Gly Pro Gly Val Glu Ala
                845                 850                 855
Phe His Val Asp Met Asp Ser Gly Leu Val Thr Thr Gln Arg Pro
                860                 865                 870
Leu Gln Ser Tyr Glu Lys Phe Ser Leu Thr Val Val Ala Thr Asp
                875                 880                 885
Gly Gly Glu Pro Pro Leu Trp Gly Thr Thr Met Leu Leu Val Glu
                890                 895                 900
Val Ile Asp Val Asn Asp Asn Arg Pro Val Phe Val Arg Pro Pro
                905                 910                 915
Asn Gly Thr Ile Leu His Ile Arg Glu Glu Ile Pro Leu Arg Ser
                920                 925                 930
Asn Val Tyr Glu Val Tyr Ala Thr Asp Lys Asp Glu Gly Leu Asn
                935                 940                 945
Gly Ala Val Arg Tyr Ser Phe Leu Lys Thr Ala Gly Asn Arg Asp
                950                 955                 960
Trp Glu Phe Phe Ile Ile Asp Pro Ile Ser Gly Leu Ile Gln Thr
                965                 970                 975
Ala Gln Arg Leu Asp Arg Glu Ser Gln Ala Val Tyr Ser Leu Ile
                980                 985                 990
Leu Val Ala Ser Asp Leu Gly Gln Pro Val Pro Tyr Glu Thr Met
                995                1000                1005
Gln Pro Leu Gln Val Ala Leu Glu Asp Ile Asp Asp Asn Glu Pro
               1010                1015                1020
Leu Phe Val Arg Pro Pro Lys Gly Ser Pro Gln Tyr Gln Leu Leu
               1025                1030                1035
Thr Val Pro Glu His Ser Pro Arg Gly Thr Leu Val Gly Asn Val
               1040                1045                1050
Thr Gly Ala Val Asp Ala Asp Glu Gly Pro Asn Ala Ile Val Tyr
               1055                1060                1065
```

Tyr Phe Ile Ala Ala Gly Asn Glu Glu Lys Asn Phe His Leu Gln
1070 1075 1080
Pro Asp Gly Cys Leu Leu Val Leu Arg Asp Leu Asp Arg Glu Arg
1085 1090 1095
Glu Ala Ile Phe Ser Phe Ile Val Lys Ala Ser Ser Asn Arg Ser
1100 1105 1110
Trp Thr Pro Pro Arg Gly Pro Ser Pro Thr Leu Asp Leu Val Ala
1115 1120 1125
Asp Leu Thr Leu Gln Glu Val Arg Val Val Leu Glu Asp Ile Asn
1130 1135 1140
Asp Gln Pro Pro Arg Phe Thr Lys Ala Glu Tyr Thr Ala Gly Val
1145 1150 1155
Ala Thr Asp Ala Lys Val Gly Ser Glu Leu Ile Gln Val Leu Ala
1160 1165 1170
Leu Asp Ala Asp Ile Gly Asn Asn Ser Leu Val Phe Tyr Ser Ile
1175 1180 1185
Leu Ala Ile His Tyr Phe Arg Ala Leu Ala Asn Asp Ser Glu Asp
1190 1195 1200
Val Gly Gln Val Phe Thr Met Gly Ser Met Asp Gly Ile Leu Arg
1205 1210 1215
Thr Phe Asp Leu Phe Met Ala Tyr Ser Pro Gly Tyr Phe Val Val
1220 1225 1230
Asp Ile Val Ala Arg Asp Leu Ala Gly His Asn Asp Thr Ala Ile
1235 1240 1245
Ile Gly Ile Tyr Ile Leu Arg Asp Asp Gln Arg Val Lys Ile Val
1250 1255 1260
Ile Asn Glu Ile Pro Asp Arg Val Arg Gly Phe Glu Glu Glu Phe
1265 1270 1275
Ile His Leu Leu Ser Asn Ile Thr Gly Ala Ile Val Asn Thr Asp
1280 1285 1290
Asn Val Gln Phe His Val Asp Lys Lys Gly Arg Val Asn Phe Ala
1295 1300 1305
Gln Thr Glu Leu Leu Ile His Val Val Asn Arg Asp Thr Asn Arg
1310 1315 1320
Ile Leu Asp Val Asp Arg Val Ile Gln Met Ile Asp Glu Asn Lys
1325 1330 1335
Glu Gln Leu Arg Asn Leu Phe Arg Asn Tyr Asn Val Leu Asp Val
1340 1345 1350
Gln Pro Ala Ile Ser Val Arg Leu Pro Asp Asp Met Ser Ala Leu
1355 1360 1365
Gln Met Ala Ile Ile Val Leu Ala Ile Leu Leu Phe Leu Ala Ala
1370 1375 1380
Met Leu Phe Val Leu Met Asn Trp Tyr Tyr Arg Thr Val His Lys
1385 1390 1395
Arg Lys Leu Lys Ala Ile Val Ala Gly Ser Ala Gly Asn Arg Gly
1400 1405 1410
Phe Ile Asp Ile Met Asp Met Pro Asn Thr Asn Lys Tyr Ser Phe
1415 1420 1425
Asp Gly Ala Asn Pro Val Trp Leu Asp Pro Phe Cys Arg Asn Leu
1430 1435 1440
Glu Leu Ala Ala Gln Ala Glu His Glu Asp Asp Leu Pro Glu Asn
1445 1450 1455
Leu Ser Glu Ile Ala Asp Leu Trp Asn Ser Pro Thr Arg Thr His
1460 1465 1470
Gly Thr Phe Gly Arg Glu Pro Ala Ala Val Lys Pro Asp Asp Asp
1475 1480 1485

```
Arg Tyr Leu Arg Ala Ala Ile Gln Glu Tyr Asp Asn Ile Ala Lys
                1490                1495                1500
Leu Gly Gln Ile Ile Arg Glu Gly Pro Ile Lys Leu Ile Gln Thr
                1505                1510                1515
Glu Leu Asp Glu Glu Pro Gly Asp His Ser Pro Gly Gln Gly Ser
                1520                1525                1530
Leu Arg Phe Arg His Lys Pro Pro Val Glu Leu Lys Gly Pro Asp
                1535                1540                1545
Gly Ile His Val Val His Gly Ser Thr Gly Thr Leu Leu Ala Thr
                1550                1555                1560
Asp Leu Asn Ser Leu Pro Glu Glu Asp Gln Lys Gly Leu Gly Arg
                1565                1570                1575
Ser Leu Glu Thr Leu Thr Ala Ala Glu Ala Thr Ala Phe Glu Arg
                1580                1585                1590
Asn Ala Arg Thr Glu Ser Ala Lys Ser Thr Pro Leu His Lys Leu
                1595                1600                1605
Arg Asp Val Ile Met Glu Thr Pro Leu Glu Ile Thr Glu Leu
                1610                1615
```

<210> 9
<211> 1894
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7156379CD1

<400> 9

```
Met Lys Ala Met Pro Trp Asn Trp Thr Cys Leu Leu Ser His Leu
  1               5                  10                  15
Leu Met Val Gly Met Gly Ser Ser Thr Leu Leu Thr Arg Gln Pro
                20                  25                  30
Ala Pro Leu Ser Gln Lys Gln Arg Ser Phe Val Thr Phe Arg Gly
                35                  40                  45
Glu Pro Ala Glu Gly Phe Asn His Leu Val Val Asp Glu Arg Thr
                50                  55                  60
Gly His Ile Tyr Leu Gly Ala Val Asn Arg Ile Tyr Lys Leu Ser
                65                  70                  75
Ser Asp Leu Lys Val Leu Val Thr His Glu Thr Gly Pro Asp Glu
                80                  85                  90
Asp Asn Pro Lys Cys Tyr Pro Pro Arg Ile Val Gln Thr Cys Asn
                95                 100                 105
Glu Pro Leu Thr Thr Thr Asn Asn Val Asn Lys Met Leu Leu Ile
               110                 115                 120
Asp Tyr Lys Glu Asn Arg Leu Ile Ala Cys Gly Ser Leu Tyr Gln
               125                 130                 135
Gly Ile Cys Lys Leu Leu Arg Leu Glu Asp Leu Phe Lys Leu Gly
               140                 145                 150
Glu Pro Tyr His Lys Lys Glu His Tyr Leu Ser Gly Val Asn Glu
               155                 160                 165
Ser Gly Ser Val Phe Gly Val Ile Val Ser Tyr Ser Asn Leu Asp
               170                 175                 180
Asp Lys Leu Phe Ile Ala Thr Ala Val Asp Gly Lys Pro Glu Tyr
               185                 190                 195
Phe Pro Thr Ile Ser Ser Arg Lys Leu Thr Lys Asn Ser Glu Ala
```

```
                200                 205                 210
Asp Gly Met Phe Ala Tyr Val Phe His Asp Glu Phe Val Ala Ser
                215                 220                 225
Met Ile Lys Ile Pro Ser Asp Thr Phe Thr Ile Ile Pro Asp Phe
                230                 235                 240
Asp Ile Tyr Tyr Val Tyr Gly Phe Ser Ser Gly Asn Phe Val Tyr
                245                 250                 255
Phe Leu Thr Leu Gln Pro Glu Met Val Ser Pro Pro Gly Ser Thr
                260                 265                 270
Thr Lys Glu Gln Val Tyr Thr Ser Lys Leu Val Arg Leu Cys Lys
                275                 280                 285
Glu Asp Thr Ala Phe Asn Ser Tyr Val Glu Val Pro Ile Gly Cys
                290                 295                 300
Glu Arg Ser Gly Val Glu Tyr Arg Leu Leu Gln Ala Ala Tyr Leu
                305                 310                 315
Ser Lys Ala Gly Ala Val Leu Gly Arg Thr Leu Gly Val His Pro
                320                 325                 330
Asp Asp Asp Leu Leu Phe Thr Val Phe Ser Lys Gly Gln Lys Arg
                335                 340                 345
Lys Met Lys Ser Leu Asp Glu Ser Ala Leu Cys Ile Phe Ile Leu
                350                 355                 360
Lys Gln Ile Asn Asp Arg Ile Lys Glu Arg Leu Gln Ser Cys Tyr
                365                 370                 375
Arg Gly Glu Gly Thr Leu Asp Leu Ala Trp Leu Lys Val Lys Asp
                380                 385                 390
Ile Pro Cys Ser Ser Ala Leu Leu Thr Ile Asp Asp Asn Phe Cys
                395                 400                 405
Gly Leu Asp Met Asn Ala Pro Leu Gly Val Ser Asp Met Val Arg
                410                 415                 420
Gly Ile Pro Val Phe Thr Glu Asp Arg Asp Arg Met Thr Ser Val
                425                 430                 435
Ile Ala Tyr Val Tyr Lys Asn His Ser Leu Ala Phe Val Gly Thr
                440                 445                 450
Lys Ser Gly Lys Leu Lys Lys Ile Arg Val Asp Gly Pro Arg Gly
                455                 460                 465
Asn Ala Leu Gln Tyr Glu Thr Val Gln Val Val Asp Pro Gly Pro
                470                 475                 480
Val Leu Arg Asp Met Ala Phe Ser Lys Asp His Glu Gln Leu Tyr
                485                 490                 495
Ile Met Ser Glu Arg Gln Leu Thr Arg Val Pro Val Glu Ser Cys
                500                 505                 510
Gly Gln Tyr Gln Ser Cys Gly Glu Cys Leu Gly Ser Gly Asp Pro
                515                 520                 525
His Cys Gly Trp Cys Val Leu His Asn Thr Cys Thr Arg Lys Glu
                530                 535                 540
Arg Cys Glu Arg Ser Lys Glu Pro Arg Arg Phe Ala Ser Glu Met
                545                 550                 555
Lys Gln Cys Val Arg Leu Thr Val His Pro Asn Asn Ile Ser Val
                560                 565                 570
Ser Gln Tyr Asn Val Leu Leu Val Leu Glu Thr Tyr Asn Val Pro
                575                 580                 585
Glu Leu Ser Ala Gly Val Asn Cys Thr Phe Glu Asp Leu Ser Glu
                590                 595                 600
Met Asp Gly Leu Val Val Gly Asn Gln Ile Gln Cys Tyr Ser Pro
                605                 610                 615
Ala Ala Lys Glu Val Pro Arg Ile Ile Thr Glu Asn Gly Asp His
```

```
                620                 625                 630
His Val Val Gln Leu Gln Leu Lys Ser Lys Glu Thr Gly Met Thr
                635                 640                 645
Phe Ala Ser Thr Ser Phe Val Phe Tyr Asn Cys Ser Val His Asn
                650                 655                 660
Ser Cys Leu Ser Cys Val Glu Ser Pro Tyr Arg Cys His Trp Cys
                665                 670                 675
Lys Tyr Arg His Val Cys Thr His Asp Pro Lys Thr Cys Ser Phe
                680                 685                 690
Gln Glu Gly Arg Val Lys Leu Pro Glu Asp Cys Pro Gln Leu Leu
                695                 700                 705
Arg Val Asp Lys Ile Leu Val Pro Val Glu Val Ile Lys Pro Ile
                710                 715                 720
Thr Leu Lys Ala Lys Asn Leu Pro Gln Pro Gln Ser Gly Gln Arg
                725                 730                 735
Gly Tyr Glu Cys Ile Leu Asn Ile Gln Gly Ser Glu Gln Arg Val
                740                 745                 750
Pro Ala Leu Arg Phe Asn Ser Ser Ser Val Gln Cys Gln Asn Thr
                755                 760                 765
Ser Tyr Ser Tyr Glu Gly Met Glu Ile Asn Asn Leu Pro Val Glu
                770                 775                 780
Leu Thr Val Val Trp Asn Gly His Phe Asn Ile Asp Asn Pro Ala
                785                 790                 795
Gln Asn Lys Val His Leu Tyr Lys Cys Gly Ala Met Arg Glu Ser
                800                 805                 810
Cys Gly Leu Cys Leu Lys Ala Asp Pro Asp Phe Ala Cys Gly Trp
                815                 820                 825
Cys Gln Gly Pro Gly Gln Cys Thr Leu Arg Gln His Cys Pro Ala
                830                 835                 840
Gln Glu Ser Gln Trp Leu Glu Leu Ser Gly Ala Lys Ser Lys Cys
                845                 850                 855
Thr Asn Pro Arg Ile Thr Glu Ile Ile Pro Val Thr Gly Pro Arg
                860                 865                 870
Glu Gly Gly Thr Lys Val Thr Ile Arg Gly Glu Asn Leu Gly Leu
                875                 880                 885
Glu Phe Arg Asp Ile Ala Ser His Val Lys Val Ala Gly Val Glu
                890                 895                 900
Cys Ser Pro Leu Val Asp Gly Tyr Ile Pro Ala Glu Gln Ile Val
                905                 910                 915
Cys Glu Met Gly Glu Ala Lys Pro Ser Gln His Ala Gly Phe Val
                920                 925                 930
Glu Ile Cys Val Ala Val Cys Arg Pro Glu Phe Met Ala Arg Ser
                935                 940                 945
Ser Gln Leu Tyr Tyr Phe Met Thr Leu Thr Leu Ser Asp Leu Lys
                950                 955                 960
Pro Ser Arg Gly Pro Met Ser Gly Gly Thr Gln Val Thr Ile Thr
                965                 970                 975
Gly Thr Asn Leu Asn Ala Gly Ser Asn Val Val Val Met Phe Gly
                980                 985                 990
Lys Gln Pro Cys Leu Phe His Arg Arg Ser Pro Ser Tyr Ile Val
                995                1000                1005
Cys Asn Thr Thr Ser Ser Asp Glu Val Leu Glu Met Lys Val Ser
               1010                1015                1020
Val Gln Val Asp Arg Ala Lys Ile His Gln Asp Leu Val Phe Gln
               1025                1030                1035
Tyr Val Glu Asp Pro Thr Ile Val Arg Ile Glu Pro Glu Trp Ser
```

```
                1040                1045                1050
Ile Val Ser Gly Asn Thr Pro Ile Ala Val Trp Gly Thr His Leu
                1055                1060                1065
Asp Leu Ile Gln Asn Pro Gln Ile Arg Ala Lys His Gly Gly Lys
                1070                1075                1080
Glu His Ile Asn Ile Cys Glu Val Leu Asn Ala Thr Glu Met Thr
                1085                1090                1095
Cys Gln Ala Pro Ala Leu Ala Leu Gly Pro Asp His Gln Ser Asp
                1100                1105                1110
Leu Thr Glu Arg Pro Glu Glu Phe Gly Phe Ile Leu Asp Asn Val
                1115                1120                1125
Gln Ser Leu Leu Ile Leu Asn Lys Thr Asn Phe Thr Tyr Tyr Pro
                1130                1135                1140
Asn Pro Val Phe Glu Ala Phe Gly Pro Ser Gly Ile Leu Glu Leu
                1145                1150                1155
Lys Pro Gly Thr Pro Ile Ile Leu Lys Gly Lys Asn Leu Ile Pro
                1160                1165                1170
Pro Val Ala Gly Gly Asn Val Lys Leu Asn Tyr Thr Val Leu Val
                1175                1180                1185
Gly Glu Lys Pro Cys Thr Val Thr Val Ser Asp Val Gln Leu Leu
                1190                1195                1200
Cys Glu Ser Pro Asn Leu Ile Gly Arg His Lys Val Met Ala Arg
                1205                1210                1215
Val Gly Gly Met Glu Tyr Ser Pro Gly Met Val Tyr Ile Ala Pro
                1220                1225                1230
Asp Ser Pro Leu Ser Leu Pro Ala Ile Val Ser Ile Ala Val Ala
                1235                1240                1245
Gly Gly Leu Leu Ile Ile Phe Ile Val Ala Val Leu Ile Ala Tyr
                1250                1255                1260
Lys Arg Lys Ser Arg Glu Ser Asp Leu Thr Leu Lys Arg Leu Gln
                1265                1270                1275
Met Gln Met Asp Asn Leu Glu Ser Arg Val Ala Leu Glu Cys Lys
                1280                1285                1290
Glu Ala Phe Ala Glu Leu Gln Thr Asp Ile His Glu Leu Thr Ser
                1295                1300                1305
Asp Leu Asp Gly Ala Gly Ile Pro Phe Leu Asp Tyr Arg Thr Tyr
                1310                1315                1320
Thr Met Arg Val Leu Phe Pro Gly Ile Glu Asp His Pro Val Leu
                1325                1330                1335
Arg Asp Leu Glu Val Pro Gly Tyr Arg Gln Glu Arg Val Glu Lys
                1340                1345                1350
Gly Leu Lys Leu Phe Ala Gln Leu Ile Asn Asn Lys Val Phe Leu
                1355                1360                1365
Leu Ser Phe Ile Arg Thr Leu Glu Ser Gln Arg Ser Phe Ser Met
                1370                1375                1380
Arg Asp Arg Gly Asn Val Ala Ser Leu Ile Met Thr Val Leu Gln
                1385                1390                1395
Ser Lys Leu Glu Tyr Ala Thr Asp Val Leu Lys Gln Leu Leu Ala
                1400                1405                1410
Asp Leu Ile Asp Lys Asn Leu Glu Ser Lys Asn His Pro Lys Leu
                1415                1420                1425
Leu Leu Arg Arg Thr Glu Ser Val Ala Glu Lys Met Leu Thr Asn
                1430                1435                1440
Trp Phe Thr Phe Leu Leu Tyr Lys Phe Leu Lys Glu Cys Ala Gly
                1445                1450                1455
Glu Pro Leu Phe Ser Leu Phe Cys Ala Ile Lys Gln Gln Met Glu
```

```
                1460                1465                1470
Lys Gly Pro Ile Asp Ala Ile Thr Gly Glu Ala Arg Tyr Ser Leu
                1475                1480                1485
Ser Glu Asp Lys Leu Ile Arg Gln Gln Ile Asp Tyr Lys Thr Leu
                1490                1495                1500
Val Leu Ser Cys Val Ser Pro Asp Asn Ala Asn Ser Pro Glu Val
                1505                1510                1515
Pro Val Lys Ile Leu Asn Cys Asp Thr Ile Thr Gln Val Lys Glu
                1520                1525                1530
Lys Ile Leu Asp Ala Ile Phe Lys Asn Val Pro Cys Ser His Arg
                1535                1540                1545
Pro Lys Ala Ala Asp Met Asp Leu Glu Trp Arg Gln Gly Ser Gly
                1550                1555                1560
Ala Arg Met Ile Leu Gln Asp Glu Asp Ile Thr Thr Lys Ile Glu
                1565                1570                1575
Asn Asp Trp Lys Arg Leu Asn Thr Leu Ala His Tyr Gln Val Pro
                1580                1585                1590
Asp Gly Ser Val Val Ala Leu Val Ser Lys Gln Val Thr Ala Tyr
                1595                1600                1605
Asn Ala Val Asn Asn Ser Thr Val Ser Arg Thr Ser Ala Ser Lys
                1610                1615                1620
Tyr Glu Asn Met Ile Arg Tyr Thr Gly Ser Pro Asp Ser Leu Arg
                1625                1630                1635
Ser Arg Thr Pro Met Ile Thr Pro Asp Leu Glu Ser Gly Val Lys
                1640                1645                1650
Met Trp His Leu Val Lys Asn His Glu His Gly Asp Gln Lys Glu
                1655                1660                1665
Gly Asp Arg Gly Ser Lys Met Val Ser Glu Ile Tyr Leu Thr Arg
                1670                1675                1680
Leu Leu Ala Thr Lys Gly Thr Leu Gln Lys Phe Val Asp Asp Leu
                1685                1690                1695
Phe Glu Thr Ile Phe Ser Thr Ala His Arg Gly Ser Ala Leu Pro
                1700                1705                1710
Leu Ala Ile Lys Tyr Met Phe Asp Phe Leu Asp Glu Gln Ala Asp
                1715                1720                1725
Lys His Gly Ile His Asp Pro His Val Arg His Thr Trp Lys Ser
                1730                1735                1740
Asn Cys Leu Pro Leu Arg Phe Trp Val Asn Met Ile Lys Asn Pro
                1745                1750                1755
Gln Phe Val Phe Asp Ile His Lys Asn Ser Ile Thr Asp Ala Cys
                1760                1765                1770
Leu Ser Val Val Ala Gln Thr Phe Met Asp Ser Cys Ser Thr Ser
                1775                1780                1785
Glu His Arg Leu Gly Lys Asp Ser Pro Ser Asn Lys Leu Leu Tyr
                1790                1795                1800
Ala Lys Asp Ile Pro Ser Tyr Lys Asn Trp Val Glu Arg Tyr Tyr
                1805                1810                1815
Ser Asp Ile Gly Lys Met Pro Ala Ile Ser Asp Gln Asp Met Asn
                1820                1825                1830
Ala Tyr Leu Ala Glu Gln Ser Arg Met His Met Asn Glu Phe Asn
                1835                1840                1845
Thr Met Ser Ala Leu Ser Glu Ile Phe Ser Tyr Val Gly Lys Tyr
                1850                1855                1860
Ser Glu Glu Ile Leu Gly Pro Leu Asp His Asp Asp Gln Cys Gly
                1865                1870                1875
Lys Gln Lys Leu Ala Tyr Lys Leu Glu Gln Val Ile Thr Leu Met
```

1880 1885 1890
Ser Leu Asp Ser

<210> 10
<211> 326
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473626CD1

<400> 10
Met Pro Gly Thr Val Arg Arg Trp Asn Tyr Pro Pro Pro Leu Cys
1 5 10 15
Ile Ala Gln Cys Gly Gly Thr Val Glu Glu Met Glu Gly Val Ile
20 25 30
Leu Ser Pro Gly Phe Pro Gly Asn Tyr Pro Ser Asn Met Asp Cys
35 40 45
Ser Trp Lys Ile Ala Leu Pro Val Gly Phe Gly Ala His Ile Gln
50 55 60
Phe Leu Asn Phe Ser Thr Glu Pro Asn His Asp Tyr Ile Glu Ile
65 70 75
Arg Asn Gly Pro Tyr Glu Thr Ser Arg Met Met Gly Arg Phe Ser
80 85 90
Gly Ser Glu Leu Pro Ser Ser Leu Leu Ser Thr Ser His Glu Thr
95 100 105
Thr Val Tyr Phe His Ser Asp His Ser Gln Asn Arg Pro Gly Phe
110 115 120
Lys Leu Glu Tyr Gln Ala Tyr Glu Leu Gln Glu Cys Pro Asp Pro
125 130 135
Glu Pro Phe Ala Asn Gly Ile Val Arg Gly Ala Gly Tyr Asn Val
140 145 150
Gly Gln Ser Val Thr Phe Glu Cys Leu Pro Gly Tyr Gln Leu Thr
155 160 165
Gly His Pro Val Leu Thr Cys Gln His Gly Thr Asn Arg Asn Trp
170 175 180
Asp His Pro Leu Pro Lys Cys Glu Val Pro Cys Gly Gly Asn Ile
185 190 195
Thr Ser Ser Asn Gly Thr Val Tyr Ser Pro Gly Phe Pro Ser Pro
200 205 210
Tyr Ser Ser Ser Gln Asp Cys Val Trp Leu Ile Thr Val Pro Ile
215 220 225
Gly His Gly Val Arg Leu Asn Leu Ser Leu Leu Gln Thr Glu Pro
230 235 240
Ser Gly Asp Phe Ile Thr Ile Trp Asp Gly Pro Gln Gln Thr Ala
245 250 255
Pro Arg Leu Gly Val Phe Thr Arg Ser Met Ala Lys Lys Thr Val
260 265 270
Gln Ser Ser Ser Asn Gln Val Leu Leu Lys Phe His Arg Asp Ala
275 280 285
Ala Thr Gly Gly Ile Phe Ala Ile Ala Phe Ser Gly Gln Tyr Gly
290 295 300
Ser Leu Ala Trp Trp Glu Gly Pro Gly Phe Gln Val Lys Ala Glu
305 310 315

Leu Asp Ser Arg Leu His His Leu Arg Ile Met
320 325

<210> 11
<211> 2052
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4350981CB1

<400> 11
atggacggtg aggcagtccg cttctgcaca gataaccagt gtgtctccct gcacccccaa 60
gaggtggact ctgtggcaat ggctcctgca gcccccaaga taccgaggct cgttcaggct 120
accccggcat ttatggctgt gaccttggtc ttctctcttg tgactctctt tgtagtggat 180
catcaccact ttggcaggga ggcagaaatg cgagagctta tccagacatt taaaggccac 240
atggagaatt ccagtgcctg ggtagtagaa atccagatgt tgaagtgcag agtggacaat 300
gtcaattcgc agctccaggt gctcggtgat catctgggaa acaccaatgc tgacatccag 360
atggtaaaag gagttctaaa ggatgccact acattgagtt tgcagacaca gatgttaagg 420
agttccctgg agggaaccaa tgctgagatc cagaggctca aggaagacct tgaaaaggca 480
gatgctttaa ctttccagac gctgaatttc ttaaaaagca gtttagaaaa caccagcatt 540
gagctccacg tgctaagcag aggcttagaa aatgcaaact ctgaaattca gatgttgaat 600
gccagtttgg aaacggcaaa tacccaggct cagttagcca atagcagttt aaagaacgct 660
aatgctgaga tctatgtttt gagaggccat ctagatagtg tcaatgactt gaggacccag 720
aaccaggttt taagaaatag tttggaagga gccaatgctg agatccaggg actaaaggaa 780
aatttgcaga acacaaatgc tttaaactcc cagacccagg cctttataaa aagcagtttt 840
gacaacacta gtgctgagat ccagttctta agaggtcatt tggaaagagc tggtgatgaa 900
attcacgtgt taaaaaggga tttgaaaatg gtcacagccc agacccaaaa agcaaatggc 960
cgtctggacc agacagatac tcagattcag gtattcaagt cagagatgga aaatgtgaat 1020
accttaaatg cccagattca ggtcttaaat ggtcatatga aaaatgccag cagagagata 1080
cagaccctaa aacaaggaat gaagaatgct tcagccttaa cttcccagac ccagatgtta 1140
gacagcaatc tgcagaaggc cagtgccgag atccagaggt taagagggga tctagagaac 1200
accaaagctc taaccatgga aatccagcag gagcagagtc gcctgaagac cctccatgtg 1260
gtcattactt cacaggaaca gctacaaaga acccaaagtc agcttctcca gatggtcctg 1320
caaggctgga agttcaatgg tggaagctta tattattttt ctagtgtcaa gaagtcttgg 1380
catgaggctg agcagttctg cgtgtcccag ggagcccatc tggcatctgt ggcctccaag 1440
gaggagcagg catttctggt agagttcaca agtaaagtgt actactggat cggtctcact 1500
gacaggggca cagagggctc ctggcgctgg acagatggga caccattcaa cgccgcccag 1560
aacaaagcat aaaaaatctc aaccgtggtt ttgccagatt gaaggcactt gtgccatttc 1620
ttccccaaag caggaagccc agcaaagttg atatccttaa aggtgcgact gaatatatac 1680
aggttctcag tgatcttttg gaaggagcca aagactcaaa gaaacaagac ccagatgagc 1740
agagctatag taacaacact tctgaatcac atacatcctc ggcaagacag ctgtcaagaa 1800
acatcaccca acatatcagc tgtgctttcg gcttgaagaa tgaagaggaa gggccttggg 1860
cagatggtgg cagtggtgag ccagcacatg cttgtcgcca cagtgtgatg tctacgactg 1920
aaattatctc cccaaccaga agtctggata gattcccaga agtagaactg ctgagtcaca 1980
gacttccaca agtatgaaaa atgaaaaggc ccagggttac cttctagaga caaataaatg 2040
cagtcttgaa aa 2052

<210> 12
<211> 4234
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature

<223> Incyte ID No: 7596315CB1

<400> 12

ggccctcact aaagggagta tgctggtgcc gcttcccacc gtcctctccc ccttactggc 60
agagcgcgct gcgggcggac tcccgggccc ggagcagccc accggccacc ccaccgccca 120
cccggctccc ggtgtctcct cccggccgct ctacccagca actttccgtg ctttgttccc 180
cgactggaaa tgctttacgg aagcgtcttg gacagggtct ccgccaggcg acaagagctc 240
ggtgctgaga tgtgttacgt tctcatctcc ccatcaatta tggatggaaa caaataagga 300
agagtcaatt ttgctgagcc ccttctccgg caacgagagg cgttctgcag ccgggaggga 360
gccgccgctc gcgccggcag ccgctggcag gggcatggtg aggaggaagg tagctcagtg 420
gcatttctga gcaggggcca ccctgacttc accttggccc accatgaggg tcttcctgct 480
ttgtgcctac atactgctgc tgatggtttc ccagttgagg gcagtcagct ttcctgaaga 540
tgatgaaccc cttaatactg tcgactatca ctattcaagg caatatccgg tttttagagg 600
acgcccttca ggcaatgaat cgcagcacag gctggacttt cagctgatgt tgaaaattcg 660
agacacactt tatattgctg gcagggatca agtttataca gtaaacttaa atgaaatgcc 720
caaaacagaa gtaataccaa acaagaaact gacatggcga tcaagacaac aggatcgaga 780
aaactgtgct atgaaaggca agcataaaga tgaatgccac aactttatca aagtatttgt 840
tccaagaaac gatgagatgg tttttgtttg tggtaccaat gcattcaatc ccatgtgtag 900
atactacagg ttgagtacct tagaatatga tggggaagaa attagtggcc tggcaagatg 960
cccatttgat gccagacaaa ccaatgttgc cctctttgct gatgggaagc tgtattctgc 1020
cacagtggct gacttcttgg ccagcgatgc cgttatttat cgaagcatgg gtgatggatc 1080
tgcccttcgc acaataaaat atgattccaa atggataaaa gagccacact ttcttcatgc 1140
catagaatat ggaaactatg tctatttctt ctttcgagaa atcgctgtcg aacataataa 1200
tttaggcaag gctgtgtatt cccgcgtggc ccgcatatgt aaaaacgaca tgggtggttc 1260
ccagcgggtc ctggagaaac actggacttc atttctaaag gctcggctga actgttctgt 1320
ccctggagat tcgtttttct actttgatgt tctgcagtct attacagaca taatacaaat 1380
caatggcatc cccactgtgg tcggggtgtt taccacgcag ctcaatagca tccctggttc 1440
tgctgtctgt gcatttagca tggatgacat tgaaaaagta ttcaaaggac ggtttaagga 1500
acagaaaact ccagattctg tttggacagc agttcccgaa gacaaagtgc caaagccaag 1560
gcctggctgt tgtgcaaaac acggccttgc cgaagcttat aaaacctcca tcgatttccc 1620
ggatgaaact ctgtcattca tcaaatctca tcccctgatg gactctgccg ttccacccat 1680
tgctgatgag ccctggttca caaagactcg ggtcaggtac agactgacgg ccatctcagt 1740
ggaccattca gccggaccct accagaacta cacagtcatc tttgttggct ctgaagctgg 1800
catggtactt aaagttctgg cgaagaccag tcctttctct ttgaacgaca gcgtattact 1860
ggaagagatt gaagcctaca accatgcaaa gtgcaatgct gagaatgagg aagacaaaaa 1920
ggtcatctca ttacagttgg ataaagatca ccacgcttta tatgtggcgt tctctagctg 1980
cattatccgc atcccctca gtcgctgtga gcgttatgga tcatgtaaaa agtcttgtat 2040
tgcatctcgt gacccgtatt gtggctggtt aagccaggga tcctgtggta gagtgacccc 2100
agggatgctt gctgaaggat atgaacaaga cacagaattc ggcaacacag ctcatctagg 2160
ggactgccat gaaattttgc ctacttcaac tacaccagat tacaaaatat ttggcggtcc 2220
aacatctggt gtacgatggg aagtccagtc tggagagtcc aaccagatgg tccacatgaa 2280
tgtcctcatc acctgtgtct ttgctgcttt tgttttgggg gcattcattg caggtgtggc 2340
agtatactgc tatcgagaca tgtttgttcg gaaaaacaga aagatccata aagatgcaga 2400
gtccgcccag tcatgcacag actccagtgg aagttttgcc aaactgaatg gtctctttga 2460
cagccctgtc aaggaatacc aacagaatat tgattctcct aaactgtata gtaacctgct 2520
aaccagtcgg aaagagctac cacccaatgg agatactaaa tccatggtaa tggaccatcg 2580
agggcaacct ccagagttgg ctgctcttcc cactcctgag tctacacccg tgcttcacca 2640
gaagaccctg caggccatga agagccactc agaaaaggcc catggccatg gagcttcaag 2700
gaaagaaacc cctcagtttt ttccgtctag tccgccacct cattccccat taagtcatgg 2760
gcatatcccc agtgccattg ttcttccaaa tgctacccat gactacaaca cgtctttctc 2820
aaactccaat gctcacaaag ctgaaaagaa gcttcaaaac attgatcacc ctctcacaaa 2880
gtcatccagt aagagagatc accggcgttc tgttgattcc agaaataccc tcaatgatct 2940
cctgaagcat ctgaatgacc caaatagtaa ccccaaagcc atcatgggag acatccagat 3000
ggcacaccag aacttaatgc tggatcccat gggatcgatg tctgaggtcc cacctaaagt 3060
ccctaaccgg gaggcatcgc tatactcccc tccttcaact ctccccagaa atagcccaac 3120
caagcgagtg gatgtcccca ccactcctgg agtcccaatg acttctctgg aaagacaaag 3180

aggttatcac aaaaattcct cccagaggca ctctatatct gctatgccta aaaacttaaa 3240
ctcaccaaat ggtgttttgt tatccagaca gcctagtatg aaccgtggag gatatatgcc 3300
cacccccact ggggcgaagg tggactatat tcagggaaca ccagtgagtg ttcatctgca 3360
gccttccctc tccagacaga gcagctacac cagtaatggc actcttccta ggacgggact 3420
aaagaggacg ccgtccttaa aacctgacgt gccaccaaag ccttcctttg ttcctcaaac 3480
cccatctgtc agaccactga acaaatacac atactaggcc tcaagtgtgc tattcccatg 3540
tggctttatc ctgtccgtgt tgttgagagg atgatgttgt aagggtacct taaaacaaga 3600
gactcgcttg tattttaaga gaaccaagtg gccaaagaaa ctctttctaa ctttggcaac 3660
atcagaactt gccacatgta gctactgcag caaggcttct gtgtacttgc ctgaaaacaa 3720
aggaaggtgc tggtcattcc atttcttttg tttgaagcta aagagatgtg tagctcacag 3780
gggctacctt accagtataa agagctgata acagtactca gaagaatctg tgaacaaata 3840
cttgaaaatg ggttcaatgt agactgccat tatgtgtggt cttcccatta aatgtgaaca 3900
ttttaatatg tatgcattca ccttgcctct tgcacaaatg tcaaaaaaaa gatggtaata 3960
tctcaaagaa atgaacttgt agattaccaa gcagtttgct aaaaattcaa tctttgaccc 4020
aagctgtagc atttttttt catgtgtggc atctttttca tgccaccaac aaacttgttg 4080
tgtgtgtgcg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgttgtgtac 4140
ccactaggat ttgtttaggt gcccattgca tctttttgtg ctatggagtt gtttacatta 4200
agcatgaccg aacgagagac aatactattt ccca 4234

<210> 13
<211> 2200
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 71234712CB1

<400> 13
cgggtgagcg cggcgagcgg cgaccctggt gaggagcgcg gcgcgggagg cacgttcctt 60
agctccgccg cggccgtcct ccgcggctcg aggactccgc ttccttccct cccctcccct 120
gcgctccggc ctggggtctt ggcgcggggа gcggagggaa gggacgaagg aggagtaggt 180
gaaagcgggg tgaggggcgg aagggtcccg gcgcggggtg aggcgagggc tgcctcttgt 240
tctcccgccg ctgccgccgt ctcctggtcg ggtgccgcgg ccagaggcgc gcggggctgc 300
cgaggcaccc gcactatgca ggcagactgc cggccgccgc gatggcgagc cgggcggtgg 360
tgagagccag gcgctgcccg cagtgtcccc aagtccgggc cgcggccgcc gccccgcct 420
gggccgcgct ccccctctcc cgctccctcc ctcccctgctc caactcctcc tccttctcca 480
tgcctctgtt cctcctgctc ttacttgtcc tgctcctgct gctcgaggac gctggagccc 540
agcaaggtga tggatgtgga cacactgtac taggccctga gagtggaacc cttacatcca 600
taaactaccc acagacctat cccaacagca ctgtttgtga atgggagatc cgtgtaaaga 660
tgggagagag agttcgcatc aaatttggtg actttgacat tgaagattct gattcttgtc 720
actttaatta cttgagaatt tataatggaa ttggagtcag cagaactgaa ataggcaaat 780
actgtggtct ggggttgcaa atgaaccatt caattgaatc aaaaggcaat gaaatcacat 840
tgctgttcat gagtggaatc catgtttctg gacgcggatt tttggcctca tactctgtta 900
tagataaaca agatctaatt acttgtttgg acactgcatc caatttttg gaacctgagt 960
tcagtaagta ctgcccagct ggttgtctgc ttccttttgc tgagatatct ggaacaattc 1020
ctcatggata tagagattcc tcgccattgt gcatggctgg tgtgcatgca ggagtagtgt 1080
caaacacgtt gggcggccaa atcagtgttg taattagtaa aggtattccc tattatgaaa 1140
gttctttggc taacaacgtc acatctgtgg tgggacactt atctacaagt ctttttacat 1200
ttaagacaag tggatgttat ggaacactgg ggatggagtc tggtgtgatc gcggatcctc 1260
aaataacagc atcatctgtg ctggagtgga ctgaccacac agggcaagag aacagttgga 1320
aacccaaaaa agccaggctg aaaaaacctg gaccgccttg ggctgctttt gccactgatg 1380
aataccagtg gttacaaata gatttgaata aggaaaagaa aataacaggc attataacca 1440
ctggatccac catggtggag cacaattact atgtgtctgc ctacagaatc ctgtacagtg 1500
atgatgggca gaaatggact gtgtacagag agcctggtgt ggagcaagat aagatatttc 1560
aaggaaacaa agattatcac caggatgtgc gtaataactt tttgccacca attattgcac 1620

gttttattag agtgaatcct acccaatggc agcagaaaat tgccatgaaa atggagctgc 1680
tcggatgtca gtttattcct aaaggtcgtc ctccaaaact tactcaacct ccacctcctc 1740
ggaacagcaa tgacctcaaa aacactacag cccctccaaa aatagccaaa ggtcgtgccc 1800
caaaatttac gcaaccacta caacctcgca gtagcaatga atttcctgca cagacagaac 1860
aaacaactgc cagtcctgat atcagaaata ctaccgtaac tccaaatgta accaaagatg 1920
tagcgctggc tgcagttctt gtccctgtgc tggtcatggt cctcactact ctcattctca 1980
tattagtgtg tgcttggcac tggagaaaca ggttagtaca taactagttc acctgagtcc 2040
aaaactacca aatgtgaagt agaagctaaa tatagaagat gaaaatgttt acctgtttga 2100
gagtgagagt taaggtaatt attaaaatga aaatttcatg cttctccttt attcccatta 2160
aaaataaata agttcaattc cacaatcaaa aaaaaaaaaa 2200

<210> 14
<211> 1647
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 079370CB1

<400> 14
ggacggctgc cgcatcgctg ggacaaactc ggcagcggag gcaaagttat ttcccctccc 60
aggcagcggg attccgactg gcaagatggt gcccagctct ccgcgcgcgc tcttccttct 120
gctcctgatc ctcgcctgcc ccgagccgcg ggcttcccag aactgtctca gcaaacagca 180
gctcctctcg gccatccgcc agctgcagca gctgctgaag ggccaggaga cacgcttcgc 240
cgagggcatc cgccacatga agagccggct ggccgcgctg cagaactctg tgggcagggt 300
gggcccagat gcccttccag tttcctgccc ggctctgaac acccccgcag acggcagaaa 360
gtttggaagc aagtacttag tggatcacga agtccatttt acctgcaacc ctgggttccg 420
gctggtcggg cccagcagcg tggtgtgtct tcccaatggc acctggacag gggagcagcc 480
ccactgtaga ggtatcagtg aatgctccag ccagccttgt caaaatggtg gtacatgtgt 540
agaaggagtc aaccagtaca gatgcatttg tcctccagga aggactggga accgctgtca 600
gcatcaggcc cagactgccg cccccgaggg cagcgtggcc ggcgactccg ccttcagccg 660
cgcgccgcgc tgtgcgcagg tggagcgggc tcagcactgc agctgcgagg ccggattcca 720
cctgagcggc gccgccggcg acagcgtctg ccaggacgtg aacgagtgtg agctctacgg 780
gcaggagggg cgccccccggc tctgcatgca cgcctgcgtg aacaccccgg gctcttaccg 840
ttgcacctgc cccggtggat accgaactct ggctgacggg aagagctgtg aggatgtgga 900
tgaatgtgtg ggcctgcagc cggtgtgccc ccaggggacc acatgcatca acaccggtgg 960
aagcttccag tgtgtcagcc ctgagtgccc cgagggcagc ggcaatgtga gctacgtgaa 1020
gacgtctcca ttccagtgtg agcggaaccc ctgccccatg gacagcaggc cctgccgcca 1080
tctgcccaag accatctcct tccattacct ctctctgcct tccaacctga agacgcccat 1140
cacgctcttc cgcatggcca cagcctctgc ccccggccga gctgggccca acagcctgcg 1200
gtttgggatc gtgggtggga acagccgcgg ccactttgtg atgcagcgtt cagaccggca 1260
gactggggat ctgatccttg tgcagaacct ggaggggcct cagacgctgg aggtggacgt 1320
cgacatgtcg gaatacctgg accgctcctt ccaggccaac cacgtgtcca aggtcaccat 1380
ctttgtatcc ccctatgact tctgagggta cacaggggca ctggggtgtg gagagctgac 1440
ctcatttctc ttccccgaag gctcagcttc gggcaccgac tgcgtggagc ctcccgcctg 1500
ttcccgccca ctcaccagtg cacccaggct tctagggcag cgttgcacgg cgccccatgg 1560
aatagcacgg aagagcagcc acaaaactca actgctgcca tcactctttt ttttttctg 1620
ctttgaggcc cttcccttag attatgc 1647

<210> 15
<211> 4456
<212> DNA
<213> Homo sapiens

<220>

<221> misc_feature
<223> Incyte ID No: 2496174CB1

<400> 15
ccgccggtgc gctccgcccg ggagtcggga gccgcgggga accgggcacc tgcacccgcc 60
tctgggaggt cttctcccct gtctgcctcc cggagctagg actgcagagg ggcctatcat 120
ggtgcttgca ggccccctgg ctgtctcgct gttgctgccc agcctcacac tgctggtgtc 180
ccacctctcc agctcccagg atgtctccag tgagcccagc agtgagcagc agctgtgcgc 240
ccttagcaag caccccaccg tggcctttga agacctgcag ccgtgggtct ctaacttcac 300
ctaccctgga gcccgggatt tctcccagct ggctttggac ccctccggga accagctcat 360
cgtgggagcc aggaactacc tcttcagact cagccttgcc aatgtctctc ttcttcagga 420
agatactgga gatgtgttcc accaaaataa gcgaataaat caagaaagag ggaaacatgc 480
gattcggaaa gcaggagagg aaagaaggcc acagagtggg cctccagtga ggacacgcgc 540
cgctcctgcc aaagcaaagg gaagactgag gaggagtgtc agaactacgt gcgagtcctg 600
atcgtcgccg gccggaaggt gttcatgtgt ggaaccaatg ccttttcccc catgtgcacc 660
agcagacagg tggggaacct cagccggact attgagaaga tcaatggtgt ggcccgctgc 720
ccctatgacc cacgccacaa ctccacagct gtcatctcct cccaggggga gctctatgca 780
gccacggtca tcgacttctc aggtcgggac cctgccatct accgcagcct gggcagtggg 840
ccaccgcttc gcactgccca atataactcc aagtggctta atgagccaaa cttcgtggca 900
gcctatgata ttgggctgtt tgcatacttc ttcctgcggg agaacgcagt ggagcacgac 960
tgtggacgca ccgtgtactc tcgcgtggcc cgcgtgtgca agaatgacgt ggggggccga 1020
ttcctgctgg aggacacatg gaccacattc atgaaggccc ggctcaactg ctcccgcccg 1080
ggcgaggtcc ccttctacta taacgagctg cagagtgcct tccacttgcc ggagcaggac 1140
ctcatctatg gagttttcac aaccaacgta aacagcatcg cggcttctgc tgtctgcgcc 1200
ttcaacctca gtgctatctc ccaggctttc aatggcccat ttcgctacca ggagaacccc 1260
agggctgcct ggctccccat agccaacccc atccccaatt tccagtgtgg caccctgcct 1320
gagaccggtc ccaacgagaa cctgacggag cgcagcctgc aggacgcgca gcgcctcttc 1380
ctgatgagcg aggccgtgca gccggtgaca cccgagccct gtgtcaccca ggacagcgtg 1440
cgcttctcac acctcgtggt ggacctggtg caggctaaag acacgctcta ccatgtactc 1500
tacattggca ccgagtcggg caccatcctg aaggcgctgt ccacggcgag ccgcagcctc 1560
cacggctgct acctggagga gctgcacgtg ctgccccccg ggcgccgcga gcccctgcgc 1620
agcctgcgca tcctgcacag cgcccgcgcg ctcttcgtgg ggctgagaga cggcgtcctg 1680
cgggtcccac tggagaggtg cgccgcctac cgcagccagg gggcatgcct gggggcccgg 1740
gacccgtact gtggctggga cgggaagcag caacgttgca gcacactcga ggacagctcc 1800
aacatgagcc tctggaccca gaacatcacc gcctgtcctg tgcggaatgt gacacgggat 1860
gggggcttcg gcccatggtc accatggcaa ccatgtgagc acttggatgg ggacaactca 1920
ggctcttgcc tgtgtcgagc tcgatcctgt gattcccctc gaccccgctg tgggggcctt 1980
gactgcctgg ggccagccat ccacatcgcc aactgctcca ggaatggggc gtggacccccg 2040
tggtcatcgt gggcgctgtg cagcacgtcc tgtggcatcg gcttccaggt ccgccagcga 2100
agttgcagca accctgctcc ccgccacggg ggccgcatct gcgtgggcaa gagccgggag 2160
gaacggttct gtaatgagaa cacgccttgc ccggtgccca tcttctgggc ttcctggggc 2220
tcctggagca agtgcagcag caactgtgga gggggcatgc agtcgcggcg tcgggcctgc 2280
gagaacggca actcctgcct gggctgcggc gtggagttca agacgtgcaa ccccgagggc 2340
tgccccgaag tgcggcgcaa cacccccctgg acgccgtggc tgcccgtgaa cgtgacgcag 2400
ggcggggcac ggcaggagca gcggttccgc ttcacctgcc gcgcgccccct tgcagacccg 2460
cacggcctgc agttcggcag gagaaggacc gagacgagga cctgtcccgc ggacggctcc 2520
ggctcctgcg acaccgacgc cctggtggag gtcctcctgc gcagcgggag cacctccccg 2580
cacacggtga gcgggggctg ggccgcctgg ggcccgtggt cgtcctgctc ccgggactgc 2640
gagctgggct tccgcgtccg caagagaacg tgcactaacc cggagccccg caacgggggc 2700
ctgccctgcg tgggcgatgc tgccgagtac caggactgca accccccaggc ttgcccagtt 2760
cggggtgctt ggtcctgctg gacctcatgg tctccatgct cagcttcctg tggtgggggt 2820
cactatcaac gcacccgttc ctgcaccagc cccgcacccct ccccaggtga ggacatctgt 2880
ctcgggctgc acacggagga ggcactatgt gccacacagg cctgcccaga aggctggtcg 2940
ccctggtctg agtggagtaa gtgcactgac gacggagccc agagccgaag ccggcactgt 3000
gaggagctcc tcccagggtc cagcgcctgt gctggaaaca gcagccagag ccgccctgc 3060
ccctacagcg agattcccgt catcctgcca gcctccagca tggaggaggc caccggctgt 3120

gcagggttca atctcatcca cttggtggcc acgggcatct cctgcttctt gggctctggg 3180
ctcctgaccc tagcagtgta cctgtcttgc cagcactgcc agcgtcagtc ccaggagtcc 3240
acactggtcc atcctgccac ccccaaccat ttgcactaca agggcggagg caccccgaag 3300
aatgaaaagt acacacccat ggaattcaag accctgaaca agaataactt gatccctgat 3360
gacagagcca acttctaccc attgcagcag accaatgtgt acacgactac ttactaccca 3420
agcccctga acaaacacag cttccggccc gaggcctcac ctggacaacg gtgcttcccc 3480
aacagctgat accgccgtcc tggggacttg ggcttcttgc cttcataagg cacagagcag 3540
atggagatgg gacagtggag ccagtttggt tttctccctc tgcactaggc caagaacttg 3600
ctgccttgcc tgtgggggggt cccatccggc ttcagagagc tctggctggc attgaccatg 3660
ggggaaaggg ctggtttcag gctgacatat ggccgcaggt ccagttcagc ccaggtctct 3720
catggttatc ttccaaccca ctgtcacgct gacactatgc tgccatgcct gggctgtgga 3780
cctactgggc atttgaggaa ttggagaatg gagatggcaa gagggcaggc ttttaagttt 3840
gggttggaga caacttcctg tggcccccac aagctgagtc tggccttctc cagctggccc 3900
caaaaaaggc ctttgctaca tcctgattat ctctgaaagt aatcaatcaa gtggctccag 3960
tagctctgga ttttctgcca gggctgggcc attgtggtgc tgccccagta tgacatggga 4020
ccaaggccag cgcaggttat ccacctctgc ctggaagtct atactctacc cagggcatcc 4080
ctctggtcag aggcagtgag tactgggaac tggaggctga cctgtgctta gaagtccttt 4140
aatctgggct ggtacaggcc tcagccttgc cctcaatgca cgaaaggtgg cccaggagag 4200
aggatcaatg ccacaggagg cagaagtctg gcctctgtgc ctctatggag actatcttcc 4260
agttgctgct caacagagtt gttggctgag acctgcttgg gagtctctgc tggcccttca 4320
tctgttcagg aacacacaca cacacacact cacacacgca cacacaatca caatttgcta 4380
cagcaacaaa aaagacattg ggctgtggca ttattaatta aagatgatat cccagtctcc 4440
aaaaaaaaaa aaaagg 4456

<210> 16
<211> 3574
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4097936CB1

<400> 16
atttaggtga cactatagaa gagcccagtg tgctggaaag ggaccctcgc cccgtcctcg 60
gctgtccagt cctcctcctc gcagaccccg gcggttccta ccccaggccg caggggagac 120
ggtgccccaa ggcaggcttc atatcctgaa cgctgggatc ccccaggaca ttccctggcc 180
cccaggcccc aggtcccagg ccccagggct gagctgtggg caggccccac ctggcctctg 240
caatgtcacc gcctctgtgt cccctccttc tcctggctgt gggcctgcgg ctggctggaa 300
ctctcaaccc cagtgatccc aatacctgca gcttctggga aagcttcact accaccacca 360
aggagtccca ctcccgcccc ttcagcctgc tcccctcaga gccctgcgag cggccctggg 420
agggcccccа tacttgcccc cagcccacgg ttgtataccg gaccgtgtac cgtcaggtgg 480
tgaagacgga ccaccgccag cgcctgcagt gctgccatgg cttctatgag agcagggggt 540
tctgtgtccc gctctgtgcc caggagtgtg tccatggccg ttgtgtggca cccaatcagt 600
gccaatgtgt gccaggctgg cgggggcgacg actgttccag tgagtgtgcc ccaggaatgt 660
gggggccaca gtgtgacaag ccctgcagct gcggcaacaa cagctcgtgt gatcccaaga 720
gtggggtatg ttcttgccct tctggtctgc agcccccgaa ctgccttcag ccctgtaccc 780
ctggctacta tggccctgcc tgccagttcc gctgccagtg ccatggggca ccctgcgatc 840
cccagactgg agcctgcttc tgccccgcag agagaactgg gcccagctgt gacgtgtcct 900
gttcccaggg cacttctggc ttcttctgcc ccagcaccca tccttgccaa aatggaggtg 960
tcttccaaac cccacagggc tcctgcagct gcccccctgg ctggatgggc accatctgct 1020
ccctgccctg cccagagggc tttcacggac ccaactgctc ccaggaatgt cgctgccaca 1080
acggcggcct ctgtgaccga ttcactgggc agtgccgctg cgctccgggt tacactgggg 1140
atcggtgccg ggaggagtgc ccggtgggcc gctttgggca ggactgtgct gagacgtgcg 1200
actgcgcccc ggacgcccgt tgcttcccgg ccaacggcgc atgtctgtgc gaacacggct 1260
tcactgggga ccgctgcacg gatcgcctct gccccgacgg cttctacggt ctcagctgcc 1320

aggcccccctg cacctgcgac cgggagcaca gcctcagctg ccacccgatg aacggggagt 1380
gctcctgcct gccgggctgg gcgggcctcc actgcaacga gagctgcccg caggacacgc 1440
atgggccagg gtgccaggag cactgtctct gcctgcacgg tggcgtctgc caggctacca 1500
gcggcctctg tcagtgcgcg ccgggttaca cgggccctca ctgtgctagt ctttgtcctc 1560
ctgacaccta cggtgtcaac tgttctgcac gctgctcatg tgaaaatgcc atcgcctgct 1620
cacccatcga cggcgagtgc gtctgcaagg aaggttggca gcgtggtaac tgctctgtgc 1680
cctgcccacc cggaacctgg ggcttcagtt gcaatgccag ctgccagtgt gcccatgagg 1740
cagtctgcag cccccaaact ggagcctgta cctgcacccc tgggtggcat ggggcccact 1800
gccagctgcc ctgtccgaag gggcagtttg gagaaggttg tgccagtcgc tgtgactgtg 1860
accactctga tggctgtgac cctgttcatg gacgctgtca gtgccaggct ggctggatgg 1920
gtgcccgctg ccacctgtcc tgccctgagg gcttatgggg agtcaactgt agcaacacct 1980
gcacctgcaa gaatgggggc acctgtctcc ctgagaatgg caactgcgtg tgtgcacccg 2040
gattccgggg cccctcctgc cagagatcct gtcagcctgg ccgctatggc aaacgctgtg 2100
tgccctgcaa gtgcgctaac cactccttct gccaccccta gaacgggacc tgctactgcc 2160
tggctggctg gacaggcccc gactgctccc agcgctgccc tctggggaca tttggtgcta 2220
actgctccca gccatgccag tgtggtcctg gagaaaagtg ccacccagag actggggcct 2280
gtgtatgtcc cccagggcac agtggtgcac cttgcaggat tggaatccag gagcccttta 2340
ctgtgatgcc gaccactcca gtagcgtata actcgctggg tgcagtgatt ggcattgcag 2400
tgctggggtc ccttgtggta gccctggtgg cactgttcat tggctatcgg cactggcaaa 2460
aaggcaagga gcaccaccac ctggctgtgg cttacagcag cgggcgcctg gacggctccg 2520
agtatgtcat gccagatgtc cctcccagct acagtcacta ctactccaac cccagctacc 2580
acaccctgtc gcagtgctcc ccaaaccccc cacccccctaa caaggttcca ggcccgctct 2640
ttgccagcct gcagaaacct gagcggccag gtggggccca agggcatgat aaccacacca 2700
ccctgcctgc tgactggaag caccgccggg agcccccctcc agggcctctg gacaggggga 2760
gcagccgcct ggaccgaagc tacagctata gctacagcaa tggcccaggc ccattctaca 2820
ataaagggct catctctgaa gaggagctcg gggccagtgt ggcttccctg agcagtgaga 2880
acccatatgc caccatccgg gacctgccca gcttgccagg gggcccccgg gagagcagct 2940
acatggagat gaaaggccct ccctcaggat ctccccccag gcagcctcct cagttctggg 3000
acagccagag gcggcggcaa ccccagccac agagagacag tggcacctac gagcagccca 3060
gccccctgat ccatgaccga gactctgtgg gctcccagcc ccctctgcct ccgggcctac 3120
cccccggcca ctatgactca cccaagaaca gccacatccc tggacattat gacttgcctc 3180
cagtacggca tcccccatca cctccacttc gacgccagga ccgttgagga gccaggatgg 3240
tatggcagag gccagcacac ctggctgttg ctgctcaagg ctggggacag agcctagtgt 3300
acccctgcca ggagcaggga gtggaccggc aggctgtgaa catgaacaac gcttaacaga 3360
gcaagtgatg ggagccttgt tcctgggttc taccatggga gacgctgatc agcaggatgc 3420
ctggctccct ttcccaaccc actgctccca aggcctccag ggccctgtgt acataaactg 3480
gtgggttgga agttgctggg taactctgat ttcagacatg cgtgtggggt acctttctg 3540
tgcatgctca gcctgggctc tgtgcgtgtg tgtg 3574

<210> 17
<211> 3562
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2523646CB1

<400> 17
gcctgcagct tgggcagcca aggagcattt gacagaggaa gcgaaaccac caagaaacag 60
gtttggggac cgtaagaccc ctcatccaga actctaggtt aagcaaatca tctttggaac 120
tggtgctctt tacagccacc aggaggccat caaaaaatga agaagaggaa aagaaaagat 180
accgcatttt aaattttggg gcccctccta gaatgaaaat gacaaggcct cgggtctggt 240
tggctgaagg atgtagagaa tgggctttga gggactcagc cctgatggcg cagctgctcc 300
gcactggctc acccttgtac ttgctttgct ctcatcccca gaatacacca gtggggacgc 360
ccatcttcat cgtgaatgcc acagaccccg acttgggggc agggggcagc gtcctctact 420

ccttccagcc ccctcccaa ttcttcgcca ttgacagcgc ccgcggtatc gtcacagtga 480
tccgggagct ggactacgag accacacagg cctaccagct cacggtcaac gccacagatc 540
aagacaagac caggcctctg tccaccctgg ccaacttggc catcatcatc acagatgtcc 600
aggacatgga ccccatcttc atcaacctgc cttacagcac caacatctac gagcattctc 660
ctccgggcac gacggtgcgc atcatcaccg ccatagacca ggataaagga cgtccccggg 720
gcattggcta caccatcgtt tcagggaata ccaacagcat ctttgccctg gactacatca 780
gcggagtgct gaccttgaat ggcctgctgg accgggagaa ccccctgtac agccatggct 840
tcatcctgac tgtgaagggc acggagctga acgatgaccg caccccatct gacgctacag 900
tcaccacgac cttcaatatc ctggttattg acatcaatga caatgccccg gagttcaaca 960
gctccgagta cagcgtggcc atcactgagc tggcacaggt cggctttgcc cttccactct 1020
tcatccaggt ggtggacaag gatgagaatt tgggcctgaa cagcatgttt gaggtgtact 1080
tggtggggaa caactcccac cacttcatca tctccccgac ctccgtccag gggaaggcgg 1140
acattcgtat tcgggtggcc atcccactgg actacgagac cgtggaccgc tacgactttg 1200
atctctttgc caatgagagt gtgcctgacc atgtgggcta tgccaaggtg aagatcactc 1260
tcatcaatga aaatgacaac cggcccatct tcagccagcc actgtacaac atcagcctgt 1320
acgagaacgt caccgtgggg acctctgtgc tgacagtcct ggcaactgac aatgatgcag 1380
gcacctttgg ggaagtcagc tacttcttca gtgatgaccc tgacaggttc tcgctggaca 1440
aggacacggg actcatcatg ctgattgcca ggctggacta tgagctcatc cagcgcttca 1500
ccctgacgat cattgcccgg gacgggggcg gcgaggagac cacaggccgg gtcaggatca 1560
atgtgttgga tgtcaacgac aacgtgccca ccttccagaa ggatgcctac gtgggtgctc 1620
tgcgggagaa cgagccttct gtcacacagc tggtgcggct ccgggcaaca gatgaagact 1680
cccctcccaa caaccagatc acctacagca ttgtcagtgc atctgccttt ggcagctact 1740
tcgacatcag cctgtacgag ggctatggag tgatcagcgt cagtcgcccc ctggattatg 1800
aacagatatc caatgggctg atttatctga cggtcatggc catggatgct ggcaaccccc 1860
ctctcaacag caccgtccct gtcaccatcg aggtgtttga tgagaatgac aaccctccca 1920
ccttcagcaa gcccgcctac ttcgtctccg tggtggagaa catcatggca ggagccacgg 1980
tgctgttcct gaatgccaca gacctggacc gctcccggga gtacggccag gagtccatca 2040
tctactcctt ggaaggctcc acccagtttc ggatcaatgc ccgctcaggg gaaatcacca 2100
ccacgtctct gcttgaccga gagaccaagt ctgaatacat cctcatcgtt cgcgcagtgg 2160
acgggggtgt gggccacaac cagaaaactg gcatcgccac cgtaaacatc accctcctgg 2220
acatcaacga caaccacccc acgtggaagg acgcacccta ctacatcaac ctggtggaga 2280
tgaccctcc agactctgac gtgaccacgg tggtggctgt tgacccagac ctgggggaga 2340
atggcaccct ggtgtacagc atccagccac ccaacaagtt ctacagcctc aacagcacca 2400
cgggcaagat ccgcaccacc cacgccatgc tggaccggga gaaccccgac ccccatgagg 2460
ccgagctgat gcgcaaaatc gtcgtctctg ttactgactg tggcaggccc cctctgaaag 2520
ccaccagcag tgccacagtg tttgtgaacc tcttggatct caatgacaat gaccccacct 2580
ttcagaacct gcctttttgtg gccgaggtgc ttgaaggcat cccggcgggg gtctccatct 2640
accaagtggt ggccatcgac ctcgatgagg gcctgaacgg cctggtgtcc taccgcatgc 2700
cggtgggcat gccccgcatg gacttcctca tcaacagcag cagcggcgtg gtggtcacca 2760
ccaccgagct ggaccgcgag cgcatcgcgg agtaccagct gcgggtggtg gccagtgatg 2820
caggcacgcc caccaagagc tccaccagca cgctcaccat ccatgggtgc tcagagggct 2880
gcatgtggtc ctgtatggga agcactcagc atgggcttgg cacgctggac aagctcgtaa 2940
atgtgctgga tgtgaacgac gagacgccca ccttcttccc ggccgtgtac aatgtgtctg 3000
tgtccgagga cgtgccacgc gagttccggg tggtctggct gaactgcacg gacaacgacg 3060
tgggcctcaa tgcagagctc agctacttca tcacaggtgc tgccccggcc tccgcccacc 3120
tgtgcaggcc tcctggggcc ctgcctccac ccctcccaga tggacagcca gactaggtgg 3180
gggcaggtga gggtggaaaa gaggtcaggg ctctactgtt gggctttagc ctctggtggt 3240
gcctcccgag gatttgctcc tggctcttcc caagggcttt gcagctggat cactctggac 3300
tggctccctg gggacctcct gaacctgttg gttgcaggga cggggagcat ctaccaaggt 3360
tcattctaga gggaggtaag gccccatgat tcctagggag gagccctgag ccccactccc 3420
cgccccaagt ctgggtgaca gagcagtgac ttggaggaat gtggcctcat ccttccttgg 3480
ggacctgttg agaattccca cctgtttaga ggcagatggt tttgatctcc ctaaatgaaa 3540
tggttttagc tcaaaaaaaa aa 3562

<210> 18
<211> 6197

<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4099073CB1

<400> 18
tcagcgatta tgtgggaacc agagtgttca cagtccagaa ctgtgtaacc gtcggccttt 60
gtgccgcgcg ttatttcgac ctcactatgg gcggattggt gtaccggggc ccccccctcg 120
aggtccgacc ggtatcgaat agcctttatt ttcgcgaatt cctgccagcc ccccaggctc 180
ctgggcaccg ccggttggag cctgccctgc ctttgcttgt agaccaggag acaggatgaa 240
ggggaggcag aggattccta cgattccagg cacccgtagc cagggctgca gcccgtgatg 300
ttttcactct cctgggagga ggatgggtac accacacagt tggatggtgc atctttgcct 360
gtctgcacct gcagctccat ggcaccatgg accctgtgct ccgagtggtg gtgcctgatc 420
tccaccacca ggcagcagta gaggccgcta tccagcaggg tcaggttgcg catggtgatg 480
gagaagttgc catggtggtc ggaggccgac tccagcccgt ggcgctgagc caggtcgtgg 540
ctggtgttgg cagcctggtg gcctccatgg tgcaggtgaa ggtcctggaa cgtgaggttg 600
cggatgggcc ggcgctctga gcaggtctgc acctcgcccc tcgagctgcg gtaccacgtc 660
ttgtagaagg tcacatcgtg ccctttgtcc acagggccca agagcctgca ggtgagggtg 720
acgttctgcc cctcgggaca gacatacagg gaatacggcg tggcgacctt gaaggctgcc 780
accggaccta gggacgcagc caggaagaga gcgaagagca gggatcccca gcgccagctg 840
ccggcctcca gggccgtggg gacgcccatg tcgccgtcgg acgcgcagag gaacttctgg 900
tgccgggggag cgggcgggac gcggccggcg cggggaagcc tcccgcgact gagtgcgagc 960
gagtgagcgc tgcgggcggc cccagcgccg tgctccaggc acccgccccc ttcgcgcagc 1020
gccccgggg ggccgtgtgg gggaactgcc tctccgaggg ccgcgtggga ggggcttccc 1080
ggaggccggc cccgtgggcc accacgcacg tgtacgtgac cattgtggat gagaatgata 1140
acgcgcccat gttccagcag ccccactatg aggtgctgct ggatgagggc ccagacacgc 1200
tcaacaccag cctcatcacc atccaggcac tggacctgga tgagggtccc aacggcacag 1260
tcacctatgc catcgtcgca ggcaacatcg tcaacacctt ccgcatcgac agacacatgg 1320
gtgtcatcac tgctgccaaa gagctggact acgagatcag ccacggccgc tacaccctga 1380
tcgtcactgc cacagaccag tgccccatct tatcccaccg cctcacctct accaccacgg 1440
tgcttgtgaa tgtgaatgac atcaacgaca atgtgcctac cttcccccgg gactatgagg 1500
gaccatttga agtcactgag ggccagccgg ggcccagagt gtggaccttc ctggcccatg 1560
accgagactc aggacccaac gggcaggtgg agtacagcat catggatgga gaccctctgg 1620
gggagtttgt gatctctcct gtggaggggg tgctaagggt ccggaaggac gtggagctgg 1680
accgggagac catcgccttc tacaacctga ccatctgtgc ccgtgaccgg gggatgcccc 1740
cactcagctc cacaatgctg gtggggatcc gggtgctgga catcaacgac aacgaccctg 1800
tgctgctgaa cctgcccatg aacatcacca tcagcgagaa cagccctgtc tccagctttg 1860
tcgcccatgt cctggccagt gacgctgaca gtggctgcaa tgcacgcctc accttcaaca 1920
tcactgcggg caaccgcgag cgggccttct tcatcaatgc cacgacaggg atcgtcactg 1980
tgaaccggcc cctggaccgc gagcggatcc cagagtacaa gctgaccatt tctgtgaagg 2040
acaacccgga gaatccacgc atagccagga gggattatga cttgcttctg atcttccttt 2100
ctgatgagaa tgacaaccac cccctcttca ctaaaagcac ctaccaggca gaggtgatgg 2160
aaaactctcc cgctggcacc cctctcacgg tgctcaatgg gcccatcctg gccctggatg 2220
cagaccaaga catctacgcc gtggtgacct accagctgct gggtgcccag agtggcctct 2280
ttgacatcaa cagcagcacc ggattctcag tccttcaagt cacagccaca gatgaggaca 2340
gtggcctcaa tggggagctg gtctaccgaa tagaagctgg ggctcaggac cgcttcctca 2400
ttcatctggt caccggggtc atccgtgttg gtaatgccac catcgacaga gaggagcagg 2460
agtcctacag gctaacggtg gtggccaccg accggggcac cgttcctctc tcgggcacag 2520
ccattgtcac cattctgatc gatgacatca atgactcccg ccccgagttc ctcaacccca 2580
tccagacagt gagcgtgctg gagtcggctg agccaggcac tgtcattgcc aatatcacgg 2640
ccattgacca cgacctcaac ccaaagctag agtaccacat tgtcggcatt gtggccaagg 2700
acgacactga tcgcctggtg cccaaccagg aggacgcctt tgctgtgaat atcaacacag 2760
gatctgtaat ggtgaagtcc cccatgaatc gggagctggt tgccacctat gaggtcactc 2820
tctcagtgat tgacaatgcc agcgacctac cagagcgctc tgtcagtgtg ccaaatgcca 2880

agctgactgt caacgtcctg gacgtcaatg acaatacgcc ccagttcaag ccctttggga 2940
tcacctacta catggagcgg atcctggagg gggccacccc tgggaccaca ctcattgctg 3000
tggcagccgt ggaccctgac aagggcctta atgggctggt cacctacacc ctgctggacc 3060
tggtgccccc agggtatgtc cagctggagg actcctcggc agggaaggtc attgccaacc 3120
agacagtgga ctacgaggag gtgcactggc tcaactttac cgtgagggcc tcagacaacg 3180
ggtccctgcc ccgggcagct gagatccctg tctacctgga aatcgtggac atcaatgaca 3240
acaaccccat ctttgaccag ccctcctacc aggaggctgt ctttgaggat gtgcctgtgg 3300
gcacaatcat cctgacagtc actgccactg atgctgactc aggcaacttt gcactcattg 3360
agtacagcct tggagatgga gagagcaagt ttgccatcaa ccccaccacg ggtgacatct 3420
atgtgctgtc ttctctggac cgggagaaga aggaccacta tatcctgact gccttggcca 3480
aagacaaccc tggggatgta gccagcaacc gtcgcgaaaa ttcagtgcag gtggtgatcc 3540
aagtgctgga tgtcaatgac tgccggccac agttctccaa gccccagttc agcacaagcg 3600
tgtatgagaa tgagccggca ggcacctcgg tcatcaccat gatggccact gaccaggatg 3660
aaggtcccaa tggagagttg acctactcac ttgagggccc tggcgtggag gccttccatg 3720
tggacatgga ctcgggcttg gtgaccacac agcggccact gcagtcctac gagaagttca 3780
gtctgaccgt ggtggccaca gatggtggag agcccccact ctggggcacc accatgctcc 3840
tggtggaggt catcgacgtc aatgacaacc gccctgtctt tgtgcgccca cccaacggca 3900
ccatcctcca catcagagag gagatcccgc tgcgctccaa cgtgtacgag gtctacgcca 3960
cggacaagga tgagggcctc aacggggcgg tgcgctacag cttcctgaag actgcgggca 4020
accgggactg ggagttcttc atcatcgacc caatcagcgg cctcatccag actgctcagc 4080
gcctggaccg cgagtcgcag gcggtgtaca gcctcatctt ggtggccagc gacctgggcc 4140
agccagtgcc atacgagact atgcagccgc tgcaggtggc cctggaggac atcgatgaca 4200
acgaacccct tttcgtgagg cctccaaaag gcagccccca gtaccagctg ctgacagtgc 4260
ctgagcactc accacgcggc accctcgtgg gcaacgtgac aggcgcagtg gatgcagatg 4320
agggccccaa cgcgatcgtg tactacttca tcgcagccgg caacgaagag aagaacttcc 4380
atctgcagcc cgatgggtgt ctgctggtgc tgcgggacct ggaccgggag cgagaagcca 4440
tcttctcctt catcgtcaag gcctccagca atcgcagctg gacacctccc cgtggaccct 4500
ccccaaccct cgacctggtt gctgacctca cactgcagga ggtgcgcgtt gtgctagagg 4560
acatcaacga ccagccacca cgcttcacca aggctgagta cactgcaggg gtggccaccg 4620
acgccaaggt gggctcagag ttgatccagg tgctggccct ggatgcagac attggcaaca 4680
acagccttgt cttctacagc attctggcca tccactactt ccgggccctt gccaacgact 4740
ctgaagatgt gggccaggtc ttcaccatgg ggagcatgga cggcattctg cgcaccttcg 4800
acctcttcat ggcctacagc cccggctact tcgtggtgga cattgtggcc cgagacctgg 4860
caggccacaa cgacacggcc atcatcggca tctacatcct gagggacgac cagcgcgtca 4920
agatcgtcat taacgagatc cccgaccgtg tgcgcggctt cgaggaggag ttcatccacc 4980
tgctctccaa catcactggg gccattgtca atactgacaa tgtgcagttc catgtggaca 5040
agaagggccg ggtgaacttt gcgcagacag aactgcttat ccacgtggtg aaccgcgata 5100
ccaaccgcat cctggacgtg gaccgggtga tccagatgat cgatgagaac aaggagcagc 5160
tacggaatct tttccggaac tacaacgtcc tggacgtgca gcctgccatc tctgtccggc 5220
tgccggatga catgtctgcc ctgcagatgg cgatcatcgt cctggctatc ctcctgttcc 5280
tggccgccat gctctttgtc ctcatgaact ggtactacag gactgtacac aagaggaagc 5340
tcaaggccat tgtggctggc tcagctggga atcgtggctt catcgacatc atggacatgc 5400
ctaacaccaa caagtactcc tttgatggag ccaaccctgt gtggctggat cccttctgtc 5460
ggaacctgga gctggccgcc caggcggagc atgaggatga cctaccggag aacctgagtg 5520
agatcgccga cctgtggaac agccccacgc gcacccatgg aacttttggg cgtgagccag 5580
cagctgtcaa gcctgatgat gaccgatacc tgcgggctgc catccaggag tatgacaaca 5640
ttgccaagct gggccagatc attcgtgagg ggccaatcaa gctgatacag actgagctgg 5700
acgaggagcc aggagaccac agcccagggc agggtagcct gcgcttccgc cacaagccac 5760
cagtggagct caaggggccc gatgggatcc atgtggtgca cggcagcacg ggcacactgc 5820
tggccaccga cctcaacagc ctgcccgagg aagaccagaa gggcctgggc cgctcgctgg 5880
agacgctgac cgctgccgag gccactgcct tcgagcgcaa cgcccgcaca gaatccgcca 5940
aatccacacc cctgcacaaa cttcgcgacg tgatcatgga gacccccctg gagatcacag 6000
agctgtgact agacagggaa gccttgtggg tgtgagcagc acccatccac cgtcccctcc 6060
cagggagcaa gggcagggac agggccggtc gggggggacc ctccaaggcc aggccttggg 6120
gacaaccttg gcttggccct ggcagcccgc atcagctgct cagatcccac ttttgccaga 6180
cgctcattca gcatctg 6197

<210> 19
<211> 6367
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7156379CB1

<220>
<221> unsure
<222> 166
<223> a, t, c, g, or other

<400> 19
ggcccagccc acgtcccggg tcccggcatc cggcggcacg cacgggcgac atgcgccgag 60
tacgcgcgtc ccgctgcatc aggacattca gccccgggtg gacgaagggg gcaagccgcg 120
tccgcccgca gccccgagac cgccgccgct tgctcgggct ccgggnctgg ttggagaaag 180
gggtgttcgg aatcgatccc cattttccga ccttttttgtt ggacattacg cccaccttgg 240
acgccgcaag agaagctgtc agccccgcag gctctgattc ggcgccctcc gcgttcctcg 300
gctgctcccg gcttccctgt gcctcggtgg agtatttgcg ttcggggctg gggctggagg 360
aggcagccac acgcgcgcac acgcacacgt tcagaggagg gcgagaggca gcggcatagg 420
ctccatctgc agtgtcaatg cggcgctccc gctgaaggag ggaaacgcgg cgcgtccagt 480
aggggagact gcattgctga gtcctggccc tctgagggga cgactgtgcc tgagtgctgc 540
tgtgccactg ggacccgcct ctgccatgaa agccatgccc tggaactgga cctgccttct 600
ctcccacctc ctcatggtgg gcatgggctc ctccactttg ctcacccggc agccagcccc 660
gctgtcccag aagcagcggt catttgtcac attccgagga gagcccgccg agggtttcaa 720
tcacctggtg gtggatgaga ggacaggaca catttacttg ggggccgtca atcggattta 780
caagctctcc agcgacctga aggtcttggt gacgcatgag acagggccgg acgaggacaa 840
ccccaagtgt tacccacccc gcatcgtcca gacctgcaat gagcccctga ccaccaccaa 900
caatgtcaac aagatgctcc tcatagacta caaggagaac aggctgattg cctgtgggag 960
cctgtaccaa ggcatctgca agctgctgag gctggaggac ctcttcaagc tgggggagcc 1020
ttatcataag aaggagcact atctgtcagg tgtcaacgag agcggctcag tctttggagt 1080
gatcgtctcc tacagcaacc tggatgacaa gctgttcatt gccacggcag tggatgggaa 1140
gcccgagtat tttcccacca tctccagccg gaaactgacc aagaactctg aggcggatgg 1200
catgttcgcg tacgtcttcc atgatgagtt cgtggcctcg atgattaaga tcccttcgga 1260
caccttcacc atcatccctg actttgatat ctactatgtc tatggtttta gcagtggcaa 1320
ctttgtctac tttttgaccc tccaacctga gatggtgtct ccaccaggct ccaccaccaa 1380
ggagcaggtg tatacatcca agctcgtgag gctttgcaag gaggacacag ccttcaactc 1440
ctatgtagag gtgcccattg gctgtgagcg cagtggggtg gagtaccgcc tgctgcaggc 1500
tgcctacctg tccaaagcgg gggccgtgct tggcaggacc cttggagtcc atccagatga 1560
tgacctgctc ttcaccgtct tctccaaggg ccagaagcgg aaaatgaaat ccctggatga 1620
gtcggccctg tgcatcttca tcttgaagca gataaatgac cgcattaagg agcggctgca 1680
gtcttgttac cggggcgagg gcacgctgga cctggcctgg ctcaaggtga aggacatccc 1740
ctgcagcagt gcgctcttaa ccattgacga taacttctgt ggcctggaca tgaatgctcc 1800
cctgggagtg tccgacatgg tgcgtggaat tcccgtcttc acggaggaca gggaccgcat 1860
gacgtctgtc atcgcatatg tctacaagaa ccactctctg gcctttgtgg gcaccaaaag 1920
tggcaagctg aagaagatcc gggtggatgg acccaggggc aacgccctcc agtatgagac 1980
ggtgcaggtg gtggaccccg gcccagtcct ccgggatatg gccttctcca aggaccacga 2040
gcaactctac atcatgtcag agaggcagct caccagagtc cctgtggagt cctgtggtca 2100
gtatcagagc tgcggcgagt gccttggctc aggcgacccc cactgtggct ggtgtgtgct 2160
gcacaacact tgcacccgga aggagcggtg tgagcggtcc aaggagcccc gcaggtttgc 2220
ctcggagatg aagcagtgtg tccggctgac ggtccatccc aacaatatct ccgtctctca 2280
gtacaacgtg ctgctggtcc tggagacgta caatgtcccg gagctgtcag ctggcgtcaa 2340
ctgcaccttt gaggacctgt cagagatgga tgggctggtc gtgggcaatc agatccagtg 2400
ctactccccc gcagccaagg aggtgccccg gatcatcaca gagaatgggg accaccatgt 2460

cgtacagctt cagctcaaat caaaggagac cggcatgacc ttcgccagca ccagctttgt 2520
cttctacaat tgcagcgtcc acaattcgtg cctgtcctgc gtggagagtc cataccgctg 2580
ccactggtgt aaataccggc atgtctgcac ccatgacccc aagacctgct ccttccagga 2640
aggccgagtg aagctgcccg aggactgccc ccagctgctg cgagtggaca agatcctggt 2700
gcccgtggag gtgatcaagc ctatcacgct gaaggccaag aacctccccc agccccagtc 2760
tgggcagcgt ggctacgaat gcatcctcaa cattcagggc agcgagcagc gagtgcccgc 2820
cctgcgcttc aacagctcca gcgtacagtg ccagaacacc tcttattcct atgaagggat 2880
ggagatcaac aacctgcccg tggagttgac agtcgtgtgg aatgggcact tcaacattga 2940
caacccagct cagaataaag ttcacctcta caagtgtgga gccatgcgtg agagctgcgg 3000
gctgtgcctc aaggctgacc cagacttcgc atgtggctgg tgccagggcc caggccagtg 3060
caccctgcgc cagcactgcc ctgcccagga gagccagtgg ctggagctgt ctggtgccaa 3120
aagcaagtgc acaaaccccc gcatcacaga gataatcccg gtgacaggcc cccgggaagg 3180
gggcaccaag gtcactatcc gaggggagaa cctgggcctg gaatttcgcg acatcgcctc 3240
ccatgtcaag gttgctggcg tggagtgcag ccctttagtg gatggttaca tccctgcaga 3300
acagatcgtg tgtgagatgg gggaggccaa gcccagccag catgcaggct tcgtggagat 3360
ctgcgtggct gtgtgtcggc ctgaattcat ggcccggtcc tcacagctct attacttcat 3420
gacactgact ctctcagatc tgaagcccag ccgggggccc atgtccggag ggacccaagt 3480
gaccatcaca ggcaccaacc tgaatgccgg aagcaacgtg gtggtgatgt ttggaaagca 3540
gccctgtctc ttccacaggc gatctccatc ctacattgtc tgcaacacca catcctcaga 3600
tgaggtgcta gagatgaagg tgtcggtgca ggtggacagg gccaagatcc accaggacct 3660
ggtctttcag tatgtggaag accccaccat cgtgcggatt gagccagaat ggagcattgt 3720
cagtggaaac acacccatcg ccgtatgggg gacccacctg gacctcatac agaaccccca 3780
gatccgtgcc aagcatggag ggaaggagca catcaatatc tgtgaggttc tgaacgctac 3840
tgagatgacc tgtcaggcgc ccgccctcgc tctgggtcct gaccaccagt cagacctgac 3900
cgagaggccc gaggagtttg gcttcatcct ggacaacgtc cagtccctgc tcatcctcaa 3960
caagaccaac ttcacctact atcccaaccc ggtgtttgag gcctttggtc cctcaggaat 4020
cctggagctc aagcctggca cgcccatcat cctaaagggc aagaacctga tcccgcctgt 4080
ggctgggggc aacgtgaagc tgaactacac tgtgctggtt ggggagaagc cgtgcaccgt 4140
gaccgtgtca gatgtccagc tgctctgcga gtcccccaac ctcatcggca ggcacaaagt 4200
gatggcccgt gtcggtggca tggagtactc cccggggatg gtgtacattg ccccggacag 4260
cccgctcagc ctgcccgcca tcgtcagcat cgcggtggct ggcggcctcc tcatcatttt 4320
catcgtggcc gtgctcattg cctataaacg caagtcccgc gaaagtgacc tcacgctgaa 4380
gcggctgcag atgcagatgg acaacctgga gtcccgtgtg gccctggagt gcaaggaagc 4440
ctttgccgag ctgcagacgg acatccatga gctgaccagt gacctggatg gagccgggat 4500
tccgttcctg gactatagaa cttacaccat gcgggtgctg ttcccaggaa ttgaagacca 4560
ccctgtcctc cgggaccttg aggtcccggg ctaccggcag gagcgtgtgg agaaaggcct 4620
gaagctcttc gcccagctca tcaacaacaa ggtgttcctg ctgtccttca tccgcacgct 4680
tgagtcccag cgtagcttct ccatgcgcga ccgtggcaac gtggcctcac tcatcatgac 4740
cgtgctgcag agcaagctgg agtacgccac tgatgtgctg aagcagctgc tggccgacct 4800
cattgacaag aacctggaga gcaagaacca ccctaagctg ctgctcagga ggactgagtc 4860
agtggctgag aagatgctga ccaattggtt tactttcctc ctctacaagt tcctcaagga 4920
gtgtgctggg gagcccctct tctccctgtt ctgtgccatc aagcagcaga tggagaaggg 4980
ccccattgac gccatcacgg gcgaggcccg ctactccttg agcgaggaca agctcatccg 5040
ccagcagatt gactacaaaa ccctggtcct gagctgtgtc agcccagaca atgccaacag 5100
ccccgaggtc ccagtaaaga tcctcaactg tgacaccatc actcaggtca aggagaagat 5160
tctggatgcc atcttcaaga atgtgccttg ctcccaccgg cccaaagctg cagatatgga 5220
tctggagtgg cgacaaggaa gtggggcaag gatgatcttg caggatgaag acatcaccac 5280
caagattgag aatgattgga agcgactgaa cacactggcc cactaccagg tgccagatgg 5340
ttccgtggtg gcattagtgt ccaagcaggt gacagcctat aacgcagtga acaactccac 5400
cgtctccagg acctcagcaa gtaaatatga aaacatgatc cggtacacgg gcagccccga 5460
cagcctccgc tcacggacac ctatgatcac tcctgacctg gagagtggag tcaagatgtg 5520
gcacctagtg aagaaccacg agcacggaga ccagaaggag ggggaccggg ggagcaagat 5580
ggtgtctgaa atctacctga cccgactcct ggccactaag ggcacactgc agaagtttgt 5640
ggatgacctc tttgagacca tcttcagcac ggcacaccgt ggctctgccc tgcccctggc 5700
catcaagtac atgtttgact tcctggatga gcaggctgat aaacatggca ttcatgaccc 5760
gcacgtccgc catacctgga agagcaattg cctgcccctg aggttttggg tcaacatgat 5820

caagaacccg cagtttgtgt ttgacatcca taagaacagc atcacagacg cctgcctctc 5880
tgtggtggct cagaccttca tggactcttg ctccacgtca gagcaccggc tgggcaagga 5940
ctcgccctcc aacaagctgc tgtatgccaa ggacatcccc agctacaaga attgggtgga 6000
gaggtattac tcagacatag ggaagatgcc agccatcagc gaccaagaca tgaacgcata 6060
cctggctgag cagtcccgga tgcacatgaa tgagttcaac accatgagtg cactctcaga 6120
gatcttctcc tatgtgggca aatacagcga ggagatcctt ggacctctgg accacgacga 6180
ccagtgtggg aagcagaaac tggcctacaa actagaacaa gtcataaccc tcatgagctt 6240
agacagctga gaaccgtcct tccagggtcg ccctggaggg ggacacacca agccgtgcct 6300
cagtctagat tatcatcttt accaagtgca agttccgact ggcatcagca gcatcccctg 6360
agcaggg 6367

<210> 20
<211> 1615
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473626CB1

<400> 20
catttctact cagatatcag cgtatctgca gctggcttcc acttggagta caaaaatttc 60
tcactgcagg ctcaaaccct tatatcctgg cttctggcat ccctaagtgg tcactaccaa 120
atactcttca cagagaaaga tttgtggatg tgccttctgg ggaacaaaga aagaggtttc 180
tagggaaact gcttgcatac tgggactgcc tgccctaccc cctttctcca ctgggtacag 240
gaaagaatag cggtgggcct gagcagttgt ccggaacctg ctgtgcccag taacggggtg 300
aagactggcg agcgctactt ggtgaatgat gtggtgtctt tccagtgtga gccgggatat 360
gccctccagg gccacgccca catctcctgc atgcccggaa cagtgcggcg atggaactac 420
cctcctccac tctgtattgc acagtgtggg ggaacagtgg aggagatgga gggggtgatc 480
ctgagccccg gcttcccagg caactacccc agtaacatgg actgctcctg gaaaatagca 540
ctgcccgtgg gctttggagc tcacatccag ttcctgaact tctccaccga gcccaaccac 600
gactacatag aaatccggaa tggcccctat gagaccagcc gcatgatggg aagattcagt 660
ggaagcgagc ttccaagctc cctcctctcc acgtcccacg agaccaccgt gtatttccac 720
agcgaccact cccagaatcg gccaggattc aagctggagt atcaggccta tgaacttcaa 780
gagtgcccag acccagagcc ctttgccaat ggcattgtga ggggagctgg ctacaacgtg 840
ggacaatcag tgaccttcga gtgcctcccg gggtatcaat tgactggcca ccctgtcctc 900
acgtgtcaac atggcaccaa ccggaactgg gaccaccccc tgcccaagtg tgaagtccct 960
tgtggcggga acatcacttc ttccaacggc actgtgtact ccccggggtt ccctagcccg 1020
tactccagct cccaggactg tgtctggctg atcaccgtgc ccattggcca tggcgtccgc 1080
ctcaacctca gcctgctgca gacagagccc tctggagatt tcatcaccat ctgggatggg 1140
ccacagcaaa cagcaccacg gctcggcgtc ttcacccgga gcatggccaa gaaaacagtg 1200
cagagttcat ccaaccaggt cctgctcaag ttccacgtg atgcagccac agggggggatc 1260
ttcgccatag ctttctccgg tcagtatgga agcctggcct ggtgggaagg gccaggcttt 1320
caagtcaagg ctgagcttga ctcccgtctc caccatttgc ggatcatgtg accttgagtg 1380
agttgtataa cctcttggag cctcagtgtc ttcagagtta tgagaattaa atgtattagc 1440
ctatgtgaga gctctcagtg cagggttctg taaatgcaag ttttcctcct attccacact 1500
gccagggcag agaggcacag aagcccaaac cttggtgcca agtccactca ttcacatcaa 1560
ctcactggct ggatcatccc tatacctgtg ccccagctta tcccttagca ctttc 1615